Linking pattern and process in tropical rainforests

Anton J. Flügge

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of the

University College London.

Centre for Mathematics and Physics in the Life Sciences and Experimental Biology

(CoMPLEX)

University College London

April 5, 2014

To my parents without whom I could not have started and my wife without whom I could not have finished

I, Anton Jakob Flügge, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

We explore the connection between the observed spatial patterns of trees and shrubs in the rainforest and the ecological processes that shape these patterns. In particular we study the information that is captured by within-species aggregation and between species co-association. In the first part of this thesis, we introduce an individual-based model of the reproduction and death of trees of a single species. We use this as a null-model to compare within-species aggregation under idealised conditions with those found in the field. We show that the within-species aggregation of a species is expected to be strongly dependent on its local abundance. Based on this result we examine the effect of dioecy and recent changes in local abundance. We find that within-species aggregation maintains information on recent changes in local abundance. In the second part of this thesis, we examine the pair-wise cross-species spatial co-associations. In this work we do not focus on the spatial co-variation of individual species pairs, but on all pairwise combinations of a large group of species. We introduce a novel technique to normalise the cross-species co-association values that helps to make them comparable between different pairs of species, taking into account the different local abundances and within-species spatial patterns of the species. We use these normalised co-association values to find sub-communities of species that are co-located within the same spatial regions. Based on those sub-communities of species, we then investigate the effect of habitat and shade-tolerance in structuring the ecosystem.

Acknowledgements

I thank my supervisors David and Sofia for their support throughout my work on this thesis. I thank David for making sure I would not forget about the underlying biology and Sofia for asking me about the assumptions I was making, when I got caught up in creating computer models. I thank Stuart and my fellow students in journal and cake clubs for ensuring my research would not make me a hermit. I am grateful to Simon for taking the time to guide me through Yasuní's rainforest and for patiently showing me *Cecropia* flowers (while I was struggling to see anything at all through my constantly fogged glasses); and I am indebted to James for warning me of poisonous snakes and sharing his passion for nature with me. I thank the Centre for Forest Science and in particular the many people who over the past decades collected the data at Barro Colorado Island and the Yasuní forest dynamics plots. I acknowledge the grants by the National Science Foundation to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, the support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals that made the research at Barro Colorado Island possible. Of course I also thank the Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX) and their funders the EPSRC, the MRC, the British Heart Foundation, and the BBSRC, for their financial support through my PhD studentship, and for their additional support for a six-weeks field trip to the Amazonian rainforest in Ecuador. Finally, I thank my son Joshua for his telephone calls before his bedtimes in the final days of writing, motivating me to finish my thesis quickly.

Contents

List of Figures

List of Tables

List of Symbols

Symbol[a]	Description	Defined in
A	Total area of the plot.	Section 2.2.1
$A_{i,x,y}$	Area of the ring from x to y metre around individual i.	Section 2.2.1
$A_{x,y}$	Area of the ring from x to y metre around an individual.	Section 4.2.2.1
c_x	Cluster number/ID of the sub-community species x is part of.	Section 4.2.2.3
C_H	Percentage of overlap between two sub-community maps that is larger than expected for unrelated random maps.	Section 5.2.3
C_S	Percentage of overlap between two assignments of species into sub-communities that is larger than expected for random assignments of species to sub-communities.	Section 5.2.3
d	Mean dispersal distance in the individual-based model.	Section 3.2.2
$d0_x$	Sum of competition endured by individual x in the locally density dependent individual-based model.	Section 3.2.2.1
d_x	Adjusted sum of competition endured by individual x in the locally density dependent individual-based model.	Section 3.2.2.1
DBH	Diameter at breast height.	Section 2.2.3.1
$D(k)$	Sum of within cluster distances of species to their cluster centroid for a clustering with k sub-communities.	Section 4.2.2.3
f	Strength of density dependence in the locally density dependent individual-based model.	Section 3.2.2.1
H_M	Number of matching quadrants between two different sub-community maps.	Section 5.2.3
S_M	Number of matching species between two different assignments of species into sub-communities.	Section 5.2.3
H_R	Expected number of matching quadrants between two unrelated sub-community maps.	Section 5.2.3
S_R	Expected number of matching species between two unrelated random assignments of species into sub-communities.	Section 5.2.3

Continued on Next Page...

Symbol[a]	Description	Defined in
k	Number of clusters/sub-communities species are sub-divided into using k-means clustering.	Section 4.2.2.3
l_x	Propensity of individual x to die in the locally density dependent individual-based model.	Section 3.2.2.1
$m_{t,\Delta t}$	Relative change of abundance of a species between the census in year $t - \Delta t$ and year t.	Section 3.2.4
n	Size of the population and number of steps per generation in the individual-based model.	Section 2.2.2, Section 3.2.2
N	Total number of individuals of a species; abundance.	Section 2.2.1
N_t	Total number of individuals of a species at the census in year t.	Section 3.2.4
$N_{i,x,y}$	Number of conspecific individuals with a distance of between x and y metre from individual i.	Section 2.2.1
N^a	Total number of individuals of species a; abundance.	Section 4.2.2.1
$N_{x,y}^{(a_i,b)}$	Number of neighbours of species b within a distance of between x and y metre from individual i of species a.	Section 4.2.2.1
Pagel's λ	Measure of phylogenetic signal in species data.	Section 3.2.4.2
N_Q	Number of quadrants of a sub-community map.	Section 5.2.3
N_S	Number of species.	Section 5.2.3
P_a	Sum of the differences between the two normalised co-association matrices of species a, one based on the juveniles spatial distribution and the other on the adults.	Section 5.2.2.2
s	Sex ratio (percentage of females) in the dioecious species model.	Section 2.2.2
ST	Shade-tolerance index.	Section 5.2.2.2
w_i	Weight given to species i in weighted linear regression.	Equation 3.5
β_0, β_1, β_2	Regression coefficients for the linear regression models.	Section 2.2.4, Section 3.2.4
$\delta_{x,y}$	Kronecker delta.	Section 4.2.2.3
$\boldsymbol{\epsilon_1}$, $\boldsymbol{\epsilon_2}$	Vectors of independent error terms in the linear regression model.	Section 2.2.4, Section 3.2.4
$\boldsymbol{\zeta}$	Vector of indicator variables, indicating if a species is dioecious.	Section 2.2.4
$\Omega_{x,y}$	Within species relative neighbourhood density at a distance from x to y metre.	Equation 2.2.1
$\Omega_{x,y,t}$	Within species relative neighbourhood density at a distance from x to y metre at the census in year t.	Section 3.2.4
$\Omega_{x,y}^{(a,b)}$	Relative neighbourhood co-association between species a and b at a distance from x to y metre.	Section 4.2.2.1

Continued on Next Page...

Symbol[a]	Description	Defined in
$\bar{\Omega}_{x,y}^{(a,b)}$	Normalised relative neighbourhood co-association between species a and b at a distance from x to y metre.	Section 4.2.2.2
$\Omega_{x,y}^{R(a,b)}$	Relative neighbourhood co-association between species a and randomly torus translated pattern of species b at a distance from x to y metre.	Section 4.2.2.2
$\bar{\Omega}_{x,y,t}^{\text{Juveniles}(a,b)}$	Normalised relative neighbourhood co-association between the juveniles of species a and b at a distance from x to y metre in the census in year t.	Section 5.2.2.2
$\bar{\Omega}_{x,y,t}^{\text{Adults}(a,b)}$	Normalised relative neighbourhood co-association between the adults of species a and b at a distance from x to y metre in the census in year t.	Section 5.2.2.2

[a]: Symbols in bold font are vectors of multiple values, e.g. $\boldsymbol{P}$ is a vector of P_a values of a set of species A with $a \in A$.

Chapter 1

Introduction

1.1 The tropical rainforest

Tropical rainforests are among the prime biodiversity hot-spots on earth (Myers *et al.*, 2000). Many species of plants and animals and millions of insect species (Novotny *et al.*, 2002) are endemic to tropical forests and they provide a range of ecosystem services to humans (Xiao *et al.*, 2000). Forests are both threatened by climate change (Hilbert *et al.*, 2001) and their destruction can add many tons of CO_2 to the atmosphere, with the world's forest holding an amount of CO_2 that is twice the carbon content of the atmosphere (Canadell & Raupach, 2008). Yet they also help regulate the climate and act as a carbon sink (Canadell & Raupach, 2008; Tan *et al.*, 2010). The protection of rainforests is therefore an important goal for humanity. However, rising population density and land-use changes put additional pressure on this ecosystem (Wright, 2005; Hansen *et al.*, 2013), and other uses such as oil exploration and agriculture often provide more short-term economical value for the countries that harbour most of the world's tropical forests. Initiatives such as REDD (Gibbs *et al.*, 2007), eco-tourism (Gössling, 1999), and Ecudaor's Yasuní-ITT (Larrea & Warnars, 2009) aim to give forests an economic value, yet have also been criticised for setting the wrong incentives, and in many cases do not succeed in protecting large areas of forest (Kiss, 2004; Phelps *et al.*, 2010; Petherick, 2013). Researchers have tried to develop criteria to determine which areas are of most value for protection in terms of overall species richness, number of endemic species, or phylogenetic diversity or distance (Gentry, 1992; Myers *et al.*, 2000; Orme *et al.*, 2005). However, though we know much about the value of forests and the threat they are facing, many processes in tropical forests remain incompletely understood. Current research is still investigating the mechanisms that contribute to the high biodiversity of tropical forests. Monitoring forests as they undergo change and judging their ability to adapt and survive under local and global pressures is a monumental task. In order to find adaptive strategies that protect the integrity of ecosystems under human intervention and global change, an improved understanding of how forests change under external forcing and the mechanisms that maintain diversity is paramount (Figure 1.1).

In this thesis, our focus is on trees and shrubs in tropical forests (henceforth collectively

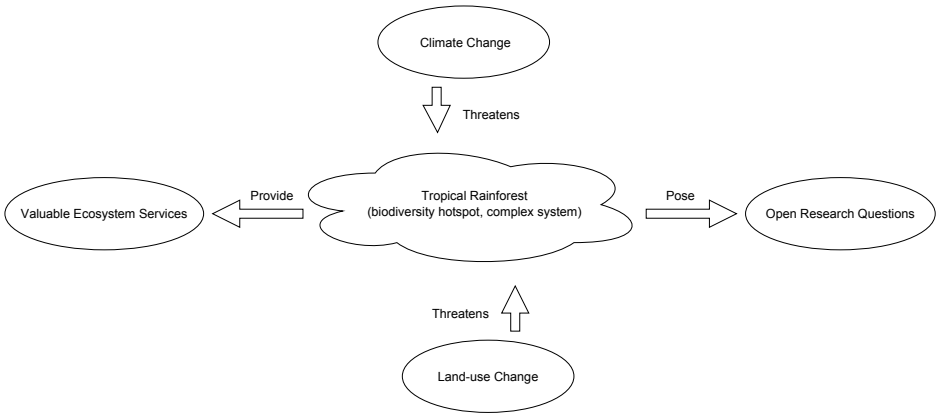


Figure 1.1: Tropical rainforests are ecosystems of high biodiversity, that provide multiple valuable services to humans. They are threatened by local pressures of land-use changes and human population increases as well as global changes of climate. In order to protect and manage rainforests, it is important to better understand the processes by which they are governed.

referred to as "trees"). Just as the rest of the biosphere of tropical rainforests, trees exhibit a high biodiversity with large numbers of different tree species even in small patches of rainforest. At the Yasuní forest research plot for example, more than 1100 species were found in an area of 50 ha (Valencia *et al.*, 2004). At the less species-rich forest on Barro Colorado Island (BCI), Panama, there are still 314 species in an area of the same size (Hubbell *et al.*, 1999), and the co-existence of such a large number of functionally similar species has long puzzled ecologists (Hutchinson, 1961). We will examine several factors that might explain this co-existence of species in detail. In particular we study the effect of habitats, light gaps and shade-tolerance (see Chapter 4 and Chapter 5), and the breeding system of species (see Chapter 2). We investigate how those factors are linked with abundance and spatial distribution of species in the forest and how temporal changes leave their mark on the forest's structure (see Chapter 3).

1.2 Spatial patterns and their link to processes

In long-lived species like trees that change only very gradually and slowly, processes are difficult to observe. It requires a lot of time to watch a tree growing, and even more time to observe forest regeneration processes over multiple generations. Even with enough time, processes such as the dispersal of seeds, habitat filtering, and competition for resources might be at least partly unobservable. Yet, all those processes leave their imprint on the current forest structure. Instead of observing a process as it happens over time, we can therefore investigate the spatial patterns that are its result. In this thesis we will focus on the spatial patterns of trees in tropical rainforests. This includes information on the spatial location of individuals together with the information on their species identity and in some analyses their stem diameter. In particular we use information from a 50 ha rainforest research plot on Barro Colorado Island (BCI; Hubbell *et al.*, 2005) for which data on spatial patterns of trees has been col-

lected since the early 1980s. We will use that data to explore mechanisms in the ecosystem that might have caused the observed pattern. Figure 1.2 shows a flow diagram on the various levels of analysis: Ecological variables like the distribution of soil chemicals (Dalling *et al.*, 2009) or the shade-tolerance of species (Valladares & Niinemets, 2008) affect ecological processes that shape the patterns and structure of the forests (Comita *et al.*, 2010; Baldeck *et al.*, 2013; Flügge *et al.*, 2013). During field work, ecologists collect data from the ecosystem which subsequently can be analysed and summarised using descriptive statistics. Those descriptions may be linked back to the ecological variables to find correlations between specific environmental conditions or species properties and an observed outcome. The main focus of this thesis is on methods to describe and analyse the collected data and on using these methods to determine how ecological processes affect the spatial pattern of trees. In particular we use summary statistics to describe the within and between species associations, i.e. whether individuals of the same species or of two different species are more or less likely to co-occur in spatial proximity to one another than they would be expected to based on a uniform random distribution throughout the forest (Wiegand & Moloney, 2004). Although we briefly touch on the topic of data collection (in Chapter 2), questions regarding the quality and reliability of the data, and the optimization of data collection for maximum information value are not the topic of this thesis. For the purpose of the work presented here, we use the data made available by the Centre for Tropical Forest Science (CTFS, 2013) from BCI (Hubbell *et al.*, 2005) and assume that it is accurate. We also do not investigate how physical properties of species or of the environment cause certain processes in the environment, we do not aim to understand the mechanistic details of biology, but take a more distanced view of the statistical relationships between different properties and observed patterns. We present methods to visualise and quantify these statistical relationships and thereby cut through the noise of random events and uncover the underlying processes that govern complex ecosystems like tropical rainforests.

Over the last couple of decades both the available data and the computational resources to analyse it have continuously increased. This allows us to use increasingly more powerful methods in our research. Mapping the spatial structure of larger area of forests needs many years of work and is beyond the scope of any single research project. Thus, early research into forest structures largely relied on small samples such as transects through the study area. Limited computer power also meant that more complex analyses were not easily possible. Early work has therefore used summary statistics for spatial patterns such as mean nearest neighbour distances that could be used on small samples and computed using limited computational resources (e.g. Clark & Evans, 1954; Hubbell, 1979). With large initiatives such as the CTFS forest network collecting data of many million trees in forests around the world (CTFS, 2013), and with growing computational means, more sophisticated measures such as Ripley's K (Ripley, 1976), the pair-correlation function (Wiegand & Moloney, 2004),

and the relative neighbourhood density Ω have become available (Condit *et al.*, 2000). These measures take into account the full spatial pattern of all individual trees to compute the aggregation of species at different scales. In addition, with increased resources, we can use computationally more demanding statistical models. In this thesis we compute the relative neighbourhood density between all pairs of species in a forest of more than 200,000 individual trees and use a computationally intensive bootstrap method to normalise our results. Bootstrapping is a way to draw random samples from the existing data to compute a measure of the variance in the data and test the significance of a result (Efron & Tibshirani, 1994; Wassermann, 2006). This can be computationally demanding, but has the advantage that we do not have to make strong assumptions about the distribution from which the data is drawn, which would be needed in more classical statistical tests. The main focus of the work presented here is to develop such new computationally intensive methods, that have now become feasible, to work on large spatially explicit multivariate ecological data sets.

1.3 Using computer models to understand complex systems

Once data has been collected, we can try to find statistical correlations between the data, or some descriptive summary statistic of it, and the ecological variables that we assume affect the processes and patterns in the ecosystem. However, it can be difficult to come up with good predictions on what relationships we expect to find and on how specific factors affect a process and ultimately the structure of the ecosystem. Simply testing all possible relationships between different environmental factors or species properties and the patterns found in nature increases the risk to obtain false positive results (Bennett *et al.*, 2009). In addition finding a correlation does not necessarily help us understand the causal mechanisms. One way to address those problems would be using controlled experiments that vary specific parts of the system in a well-defined way. However, just as observations of complex systems of long-lived species are time-consuming and effortful, this is also true for experiments. A complex ecosystem like a forest is influenced by many factors that cannot possibly all be controlled, and experiments may take multiple generations. However, both problems can also be addressed by models that capture the operational principles of the ecological processes. Models help us understand how certain processes should affect the spatial structure of a system under idealised conditions, and they can therefore be used to develop specific hypotheses of the relationships between ecological variables and patterns. Once such hypotheses have been established, statistical tests can be used to decide if there is any evidence for them in the ecosystem that is being examined.

Tropical rainforests are complex systems. The characteristic of a complex system is that its complexity arises from the interactions of its (many) components (Ottino, 2003). Individual-based models (IBMs) are a type of model that aims to simulate the behaviour of systems by simulating the behaviour of their components (Grimm *et al.*, 1999). This type of model is particularly well-suited to study complex systems, as the behaviour of each individ-

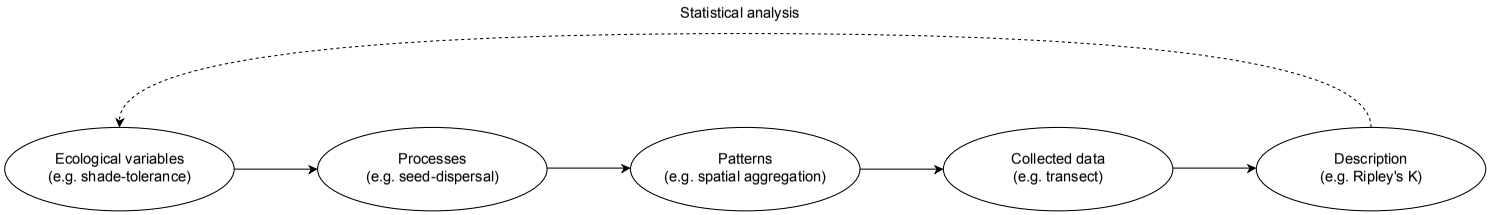


Figure 1.2: Ecological variables like species properties (e.g., shade-tolerance, drought resistance, maximum height) and environmental properties (e.g., soil nutritions, slope, altitude) affect how biological processes like the birth, growth, and death of trees operate. Those processes affect the structure of ecosystems like the spatial distribution of various species. Using data collected on an ecosystem, we can analyze its patterns by comparing suitable descriptions of the observed patterns with our knowledge of the ecological variables.

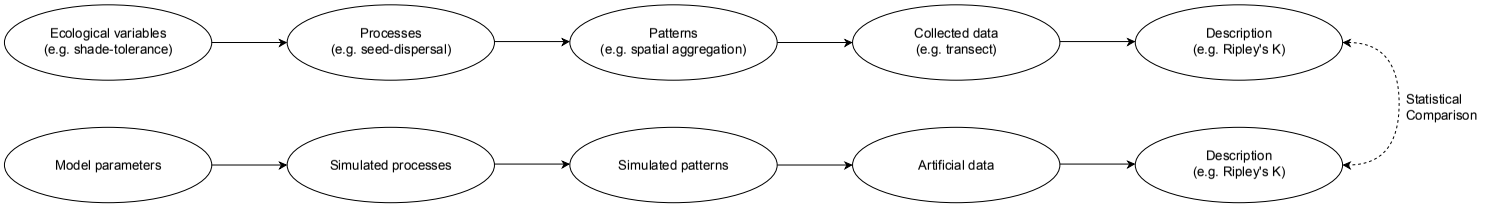


Figure 1.3: Using an individual-based model we can simulate processes found in nature. By making explicit assumptions about the interactions of the individual components or agents of a complex system we can derive the emergent pattern of the whole system. For the simulated system we can compute the same descriptions as for the natural system and make statistical comparisons between the two to determine whether they might have arisen from the same processes.

ual component of a system is often relatively simple and well-understood, but the behaviour of the system as a whole is difficult to predict as it emerges from the interactions. In contrast to statistical models that just describe the relationships between input factors and resulting patterns, IBMs make specific assumptions on the mechanisms underlying the studied processes, from which precise predictions can be derived. As Figure 1.3 shows, by creating an IBM, we simulate the whole chain from parameters that capture relevant ecological variables, the simulation of processes in the system, and the extraction of simulated patterns that are comparable to the natural system. The same methods can then be used to analyse both simulated and empirical data, and compare the results between the two. Simple IBMs have already been used before computers became available (Schelling, 1971), however, their use has increased immensely with the advance of the computer (Judson, 1994). Programming the behaviours of the components of an IBM is a simple task and computers are very good in running simulations and thereby revealing the emergent behaviour of the system that otherwise would have been difficult to describe directly. With increasing computational power, we may run increasingly complicated simulations that more closely model nature. That way, models with very precise predictions can be used to try and match the outcomes of the modelled ecosystem as closely as possible. In other fields of biology, very detailed simulations are already being used, for example for the modelling of blood flow in the heart for the purpose of predicting the effect of specific medical interventions (Quail & Taylor, 2013). However, in the field of ecology, making quantitatively precise predictions is often made difficult by insufficient knowledge and a lack of understanding of the system. We would need sufficient knowledge of the system's underlying mechanisms and current state, both of which is unlikely to be available for most complex ecological systems, such as forests. Yet, if we accept that a model will not capture all of reality, very simple IBMs can still be valuable tools to evaluate qualitatively how certain processes will affect the structure of an ecosystem. Such a simple model has the advantage that we can manipulate a specific aspect in isolation and under controlled idealised conditions. For example we can simulate the birth and death events in a population of trees and observe how the spatial aggregation of a species depends on whether individuals are hermaphroditic, i.e., combine both male and female characteristics, or dioecious with each tree being either male or female (see Chapter 2). Simple IBMs enable us to understand how specific processes are expected to influence a system which allows us to then test these hypotheses on real data. Hubbell's neutral theory of biodiversity for example is based on a very simple model of the competition of identical species (Hubbell, 2001), ignoring all the complexities of differing traits of species and the structure of the environment they are living in. Although his neutral model can therefore not be used to make specific predictions on how a particular ecosystem will change in the future, it is still useful as a model that demonstrates the surprising complexity present in an ecosystem that does not provide many different ecological niches for species. For the work presented in this thesis we have

developed an IBM where individual trees form the components of the system, and where in each step of the simulation each individual tree can either survive, die or reproduce. Using that model we make predictions on how the reproductive system of species, their abundance, and recent changes in abundance can be expected to influence the aggregation of individuals of the same species. We argue that such a model can help us formulate better hypotheses and better null-models for our analyses.

1.4 Outline of thesis

As outlined above, we investigate spatial patterns of trees in the tropical rainforests, because those spatial patterns hold information on the ecological processes that govern the system and that are themselves not easily observed. We use an IBM and novel multivariate statistical analysis tools to answer a series of ecological questions. It has been shown before that rare tree species on average have more aggregated spatial distributions (Condit *et al.*, 2000; Bleher *et al.*, 2002), yet spatial aggregation can be caused by various processes. Habitat associations (Harms *et al.*, 2001), short-range or clumped dispersal of seeds (Muller-Landau & Hardesty, 2005), and gap-dependency (Schnitzer & Carson, 2001) can all cause within-species spatial aggregation. However, it is not clear if any of those potential causes for species aggregation is responsible for the observed relationship between abundance and aggregation. Using an IBM we show that lower abundance together with spatially restricted seed dispersal alone can explain the observation that rare species are often more aggregated (Chapter 3). Based on this finding, further analyses explore particular aspects of the relationship between spatial patterns and ecological processes. It has been suggested that dioecy, in which only part of the population has the potential to produce seeds, should also lead to stronger within-species aggregation (Bleher *et al.*, 2002; Heilbuth *et al.*, 2001) and our IBM supports this expectation (Chapter 2). However, using the IBM, we also demonstrate that within-species aggregation also depends on abundance, implying that every test on the relationship of dioecy and aggregation needs to take into account the abundance of species. Using the data from BCI we show that contrary to the prediction, on average dioecious species do not show stronger within-species aggregation than hermaphrodite species (Chapter 2). Therefore, other compensatory mechanisms, such as better seed dispersal, must decrease aggregation in dioecious species. In Chapter 3 we next investigate the question of whether a static pattern can only reveal information on the processes that govern the system, or also on the temporal changes the ecosystem is undergoing. Building again on the result that rare species are expected to be more aggregated, we investigate how changes in the population size of a species affect its spatial distribution. Because the establishment of new saplings is expected to increase a species' spatial aggregation in the short-term (Law *et al.*, 2003), but at the same time a more abundant population is expected to be less aggregated at equilibrium, we expect to find a mismatch between expected and actual aggregation in species that recently experienced a change in abundance. Using the IBM we can theoretically demonstrate this effect and then test if we

find evidence supporting it in the BCI data. Finally, recognising the strong impact of abundance on spatial co-aggregation enabled us to develop a new method to normalise bi-variate co-association values between pairs of species, using a randomisation model that preserves the marginal distribution of the individual species, to make the bi-variate co-associations more comparable between different pairs of species (Chapter 4). Based on these normalised co-associations between pairs of species we explore new ways to find sub-communities of species and investigate what the effects of life-stage, habitat and shade-tolerance are in shaping the spatial structure of the forest. While previous methods often started with information on environmental conditions and then tested if species distribution followed those environmental gradients (e.g. Harms *et al.*, 2001; Kanagaraj *et al.*, 2011), we turn that process around and start from the spatial distribution of species asking if there are clusters of species that co-occur and reveal information about the structure of the environment. In Chapter 4 we prove the power of our method to detect the major habitats known to exist in the BCI forest plot, and explore the importance of shade-tolerance in structuring populations at different life-stages. In Chapter 5 we demonstrate some further extensions to the work with the normalised co-association values, suited in particular to the comparison of different data sets. With more and more high quality data sets becoming available, and simulation models being able to produce an infinite amount of artificial data sets, methods to compare the results of multiple data sets become increasingly important.

Bibliography

Baldeck, Claire A, Harms, Kyle E, Yavitt, Joseph B, John, Robert, Turner, Benjamin L, Valencia, Renato, Navarrete, Hugo, Davies, Stuart J, Chuyong, George B, Kenfack, David, *et al.* 2013. Soil resources and topography shape local tree community structure in tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, **280**(1753).

Bennett, Craig M, Miller, MB, & Wolford, GL. 2009. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Neuroimage*, **47**(1), S125.

Bleher, Bärbel, Oberrath, Reik, & Böhning-Gaese, Katrin. 2002. Seed dispersal, breeding system, tree density and the spatial pattern of trees - a simulation approach. *Basic and Applied Ecology*, **3**(2), 115–123.

Canadell, Josep G, & Raupach, Michael R. 2008. Managing forests for climate change mitigation. *science*, **320**(5882), 1456–1457.

Clark, Philip J, & Evans, Francis C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**(4), 445–453.

Comita, Liza S., Muller-Landau, Helene C., Aguilar, Salomon, & Hubbell, Stephen P. 2010. Asymmetric Density Dependence Shapes Species Abundances in a Tropical Tree Community. *Science*, **329**(5989), 330–332.

Condit, Richard, Ashton, Peter S., Baker, Patrick, Bunyavejchewin, Sarayudh, Gunatilleke, Savithri, Gunatilleke, Nimal, Hubbell, Stephen P., Foster, Robin B., Itoh, Akira, LaFrankie, James V., Lee, Hua Seng, Losos, Elizabeth, Manokaran, N., Sukumar, R., & Yamakura, Takuo. 2000. Spatial Patterns in the Distribution of Tropical Tree Species. *Science*, **288**(5470), 1414–1418.

CTFS. 2013. *Center for Tropical Forest Science*. [online] http://www.ctfs.si.edu. Accessed: 2013-06-15.

Dalling, Jim, John, Robert, Harms, Kyle, Stallard, Robert, & Yavitt, Joe. 2009. *Soil Maps of Barro Colorado Island 50 ha Plot*. `http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/soilmaps/BCIsoil.html` [Online; accessed 06-February-2013].

Efron, Bradley, & Tibshirani, Robert J. 1994. An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability).

Flügge, Anton J, Olhede, Sofia C, & Murrell, David J. 2013. Detecting sub-communities in ecosystems from multivariate spatial associations. **[under review]**.

Gentry, Alwyn H. 1992. Tropical forest biodiversity: distributional patterns and their conservational significance. *Oikos*, 19–28.

Gibbs, Holly K, Brown, Sandra, Niles, John O, & Foley, Jonathan A. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, **2**(4), 045023.

Gössling, Stefan. 1999. Ecotourism: a means to safeguard biodiversity and ecosystem functions? *Ecological Economics*, **29**(2), 303–320.

Grimm, Volker, Wyszomirski, Tomasz, Aikman, David, & Uchmański, Janusz. 1999. Individual-based modelling and ecological theory: synthesis of a workshop. *Ecological modelling*, **115**(2), 275–282.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, **342**(6160), 850–853.

Harms, Kyle E., Condit, Richard, Hubbell, Stephen P., & Foster, Robin B. 2001. Habitat Associations of Trees and Shrubs in a 50-Ha Neotropical Forest Plot. *Journal of Ecology*, **89**(6), 947–959.

Heilbuth, Jana C., Ilves, Katriina L., & Otto, Sarah P. 2001. The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. *Evolution*, **55**(5), 880–888.

Hilbert, David W, Ostendorf, Bertram, & Hopkins, Mike S. 2001. Sensitivity of tropical forests to climate change in the humid tropics of north Queensland. *Austral Ecology*, **26**(6), 590–603.

Hubbell, Stephen P. 1979. Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest. *Science*, **203**(4387), 1299–1309.

Hubbell, Stephen P. 2001. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. Princeton University Press.

Hubbell, Stephen P., Foster, Robin B., O'Brien, S. T., Harms, K. E., Condit, Richard, Wechsler, B., Wright, S. J., & de Lao, S. Loo. 1999. Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science*, **283**(5401), 554–557.

Hubbell, Stephen P., Condit, Richard, & Foster, Robin B. 2005. *Barro Colorado Forest Census Plot Data*. [online] https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Hutchinson, G. E. 1961. The Paradox of the Plankton. *The American Naturalist*, **95**(882), 137–145.

Judson, Olivia P. 1994. The rise of the individual-based model in ecology. *Trends in Ecology & Evolution*, **9**(1), 9–14.

Kanagaraj, Rajapandian, Wiegand, Thorsten, Comita, Liza S., & Huth, Andreas. 2011. Tropical tree species assemblages in topographical habitats change in time and with life stage. *Journal of Ecology*, **99**(6), 1441–1452.

Kiss, Agnes. 2004. Is community-based ecotourism a good use of biodiversity conservation funds? *Trends in Ecology & Evolution*, **19**(5), 232–237.

Larrea, Carlos, & Warnars, Lavinia. 2009. Ecuador's Yasuni-ITT Initiative: Avoiding emissions by keeping petroleum underground. *Energy for Sustainable Development*, **13**(3), 219–223.

Law, Richard, Murrell, David J., & Dieckmann, Ulf. 2003. Population growth in space and time: spatial logistic equations. *Ecology*, **84**(1), 252–262.

Muller-Landau, Helene C., & Hardesty, Britta D. 2005. Seed dispersal of woody plants in tropical forests: concepts, examples and future directions. *Pages 267–309 of:* Burslem, Pinard, & Hartley (eds), *Biotic interactions in the Tropics*. Cambridge: Cambridge University Press.

Myers, Norman, Mittermeier, Russell A, Mittermeier, Cristina G, Da Fonseca, Gustavo AB, & Kent, Jennifer. 2000. Biodiversity hotspots for conservation priorities. *Nature*, **403**(6772), 853–858.

Novotny, Vojtech, Basset, Yves, Miller, Scott E, Weiblen, George D, Bremer, Birgitta, Cizek, Lukas, & Drozd, Pavel. 2002. Low host specificity of herbivorous insects in a tropical forest. *Nature*, **416**(6883), 841–844.

Orme, C David L, Davies, Richard G, Burgess, Malcolm, Eigenbrod, Felix, Pickup, Nicola, Olson, Valerie A, Webster, Andrea J, Ding, Tzung-Su, Rasmussen, Pamela C, Ridgely, Robert S, *et al.* 2005. Global hotspots of species richness are not congruent with endemism or threat. *Nature*, **436**(7053), 1016–1019.

Ottino, Julio M. 2003. Complex systems. *AIChE Journal*, **49**(2), 292–299.

Petherick, Anna. 2013. Pipe dream. *Nature Climate Change*, **3**(10), 859–860.

Phelps, Jacob, Webb, Edward L, & Agrawal, Arun. 2010. Does REDD+ Threaten to Recentralize Forest Governance? *Science*, **328**(5976), 312–313.

Quail, Michael A, & Taylor, Andrew M. 2013. Computer Modeling to Tailor Therapy for Congenital Heart Disease. *Current cardiology reports*, **15**(9), 1–7.

Ripley, Brian D. 1976. The second-order analysis of stationary point processes. *Journal of applied probability*, 255–266.

Schelling, Thomas C. 1971. Dynamic models of segregation. *Journal of mathematical sociology*, **1**(2), 143–186.

Schnitzer, Stefan A, & Carson, Walter P. 2001. Treefall gaps and the maintenance of species diversity in a tropical forest. *Ecology*, **82**(4), 913–919.

Tan, Zhenghong, Zhang, Yiping, Yu, Guirui, Sha, Liqing, Tang, Jianwei, Deng, Xiaobao, & Song, Qinghai. 2010. Carbon balance of a primary tropical seasonal rain forest. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **115**(D4).

Valencia, R., Foster, R.B., Villa, G., Condit, R., Svenning, J.C., Hernández, C., Romoleroux, K., Losos, E., Magård, E., & Balslev, H. 2004. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *Journal of Ecology*, **92**(2), 214–229.

Valladares, Fernando, & Niinemets, Ülo. 2008. Shade tolerance, a key plant feature of complex nature and consequences. *Annual Review of Ecology, Evolution, and Systematics*, **39**(1), 237.

Wassermann, Larry. 2006. *All of nonparametric statistics*. New York: Springer.

Wiegand, T., & Moloney, K.A. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**(2), 209–229.

Wright, S Joseph. 2005. Tropical forests in a changing environment. *Trends in Ecology & Evolution*, **20**(10), 553–560.

Xiao, Han, Ouyang, Zhiyun, Zhao, Jingzhu, & Wang, Xiaoke. 2000. Forest ecosystem services and their ecological valuation–a case study of tropical forest in Jianfengling of Hainan Island]. *Ying yong sheng tai xue bao= The journal of applied ecology/Zhongguo sheng tai xue xue hui, Zhongguo ke xue yuan Shenyang ying yong sheng tai yan jiu suo zhu ban*, **11**(4), 481.

Chapter 2

Data collection, modelling, and the effect of breeding system on spatial pattern

Summary

This chapter serves two purposes. First we discuss the data available from the network of forest plots of the Centre for Tropical Forest Science (CTFS) and the methods of how this data is collected. Secondly we use a case study of a few *Cecropia* species in the Yasuní National Park, Ecuador, statistical analysis of data from Barro Colorado Island (BCI), Panama, and an individual-based computer model (IBM), to shed light on the effect a dioecious breeding system has on the spatial distribution of species. This chapter is partly informed by my own field work which I conducted at the CTFS plot in the Yasuní National Park, Ecuador. During that field trip I collected data on the spatial distribution of male and female individuals in the dioecious genus *Cecropia*. While most tree species are hermaphrodite, i.e., their flowers express both male as well as female reproductive organs, in dioecious species mature individuals have either male or female flowers. Only individuals which grow female flowers can subsequently produce seeds. The spatial distribution of species in which only part of the population has the potential to produce seeds can be expected to be different from the pattern of species in which all individuals may potentially produce seeds. We use an IBM to simulate dioecious species and compare the resulting within-species aggregation with simulated species in which all individuals may produce seeds. The model predicts dioecious species to be more aggregated than non-dioecous species of the same abundance. We test this prediction using the data from the CTFS plot at BCI and show that, contrary to this prediction, dioecious species do not show increased aggregation.

Impact of this work

In this chapter we explicitly compare the within-species aggregation of simulated species with different breeding systems and abundances. This allows us to make more specific predictions on the scale of the effect one might expect compared to previous studies. We then test for systematic differences in within-species aggregation between dioecious and non-dioecious species on a whole ecosystem level, going beyond case studies for individual species. We do not find any systematic differences between dioecious and non-dioecious species with respect to within-species aggregation which suggests that dioecious species must have mechanisms that counteract the expected increase in aggregation.

Declaration on the contributions to the work presented in this chapter:

Part of the methods used in this chapter has been published in:

The memory of spatial patterns: changes in local abundance and aggregation in a tropical forest, Anton J. Flügge, Sofia C. Olhede, and David J. Murrell, Ecology, vol. 93, nr. 7, p. 1540–1549, 2012.

The computational model presented in this chapter was developed by me. All analysis were conducted by me. The text was written by me. Sofia Olhede and David Murrell contributed by supervising and guiding my analytical work, Simon Queenborough provided training and supervision for my field work. The positions of Cecropia trees within the Yasuní forest research plot were provided by Renato Valencia and the team at the Yasuní forest research station, the additional data was collected by me. The data from BCI was provided by the CTFS.

2.1 Introduction

In this chapter we explore how being dioecious affects woody plant species in tropical rainforests and we explain why different methodological approaches are needed to tackle this question. In most shrub and tree species all individuals are able to produce seeds, either because they have bi-sexual flowers or because all individuals can produce both male and female flowers. However, there is a subset of species in which male and female flowers do not co-occur on the same individuals – these are called dioecious. While dioecy is the norm in animal species, only about 6% of angiosperm plants are dioecious (Renner & Ricklefs, 1995). Dioecy has advantages for species as it ensures out-crossing, but it also has disadvantages of which some might be particularly important for plants. When only part of the population produces offspring, and offspring are dispersed locally around the parents, such populations are expected to be more aggregated because offspring are concentrated around less parental

sources (Bleher *et al.*, 2002; Heilbuth *et al.*, 2001). This might not be important for animals with comparably wide-range movements, but could lead to significantly stronger intraspecific competition in plants (Janzen, 1970). For a plant species to be dioecious the increase in intraspecific competition has to be compensated for by an increase in fitness, or prevented by better dispersal capabilities (Heilbuth *et al.*, 2001).

Studying forests is difficult because trees are generally long-lived individuals and often surpass the life expectancy of forest ecologists. However, all processes operating in the past of an ecosystem will shape the current state of the system. Analysing the spatial patterns of trees can therefore inform us on ecologically important processes governing the system. To make the link between observed pattern and unobserved processes, we can use spatially explicit IBMs. In an IBM the behaviour of the individual components of a complex system, e.g., a tree in a forest, is modelled. Such a model can easily incorporate hypotheses on processes, such as the number of fruit a tree produces and the shape of the dispersal kernel that describes the dispersal of seeds from a parent tree. By running a simulation of the processes we are interested in, we can establish which sort of patterns should be found in a system governed by those processes. Although different processes can potentially lead to the same pattern, and IBMs therefore cannot be used to prove that any particular process has caused a given pattern, we can use them to rule out wrong hypotheses and to determine the pattern we would expect to find if a specific process would dominate the system. In this study we use an IBM of dioecious species to show that we would expect a population in which only a fraction of the individuals is able to reproduce to be more aggregated than a population in which (all else being equal) all individuals may potentially reproduce.

In order to test the hypothesis generated by the IBM, that dioecious species are expected to be more aggregated, we use data from the CTFS (2013). More specifically, we use the data from the 50 ha forest dynamic plot at BCI, Panama, in which the position and species identity of all trees and shrubs above 1 cm diameter at breast height (DBH) has been determined (Hubbell *et al.*, 2005). Although no information is available on the sex of individual trees, the breeding system, i.e., if they are dioecious or not, is known for most species (Croat, 1978). We can therefore test if there are any systematic differences in aggregation between dioecious and non-dioecious species. This may help us to decide how dioecious species compensate for the disadvantage of being dioecious.

The IBM and the statistical analysis of aggregation and breeding system can answer some questions concerning dioecy. However, to investigate if there are any differences between male and female individuals in terms of aggregation and numbers, more detailed information on the sex of each individual is needed. To explore this route we surveyed all adult *Cecropia* trees at the forest dynamic plot in Yasuní National Park, Ecuador. This allowed us to ask more detailed questions on this genus.

We suggest that only by combining all three methods, the IBM, the statistical analysis

across many species and the detailed field study of individual species, we will be able to deliver a fuller picture on complex traits such as dioecy.

2.2 Methods

2.2.1 Measure of aggregation

As a measure of the spatial aggregation of a population we use the $\Omega_{0,10}$ relative neighbourhood density as defined in Condit *et al.* (2000). This is a summary statistic of how many conspecific individuals can be found in an average neighbourhood of 10 m around an individual, standardized by the number of conspecific individuals one could expect to find for a homogeneously Poisson distributed population of the given density. $\Omega_{x,y}$ is therefore defined as:

$$\Omega_{x,y} \equiv \frac{A \sum_{i=1}^{N} N_{i,x,y}}{N \sum_{i=1}^{N} A_{i,x,y}} \tag{2.1}$$

with $N_{i,x,y}$ corresponding to the number of conspecific neighbours within the interval x to y m from a focal individual i. N is the total number of individuals of that species in the sample, $A_{i,x,y}$ the area size of the x to y m annulus around individual i and A the total area of the plot. We note that $\Omega_{x,y}$ is a version of the pair-correlation function $\hat{g}(x \pm \epsilon)$; while the pair-correlation is normally defined using one explicit parameter x for the scale at which aggregation is measured and a second implicit parameter ϵ for the width of the annulus which is used to estimate $\hat{g}(x)$, $\Omega_{x,y}$ explicitly states the inner and the outer radius of the annulus that is used to compute its value. If the inner radius x is set to zero then $\Omega_{x,y}$ is proportional to the more widely used Ripley's K statistic $\hat{K}(y)$ (Illian *et al.*, 2008).

When $\Omega_{x,y} > 1$ this indicates that the individuals are more aggregated than expected assuming there is no correlation between their positions. $\Omega_{x,y} < 1$ on the other hand means that the population is hyper-dispersed (sometimes also referred to as spatially segregated, or evenly-spaced). Even for a homogeneous Poisson process, when the expected number of individuals is small, $\Omega_{x,y}$ values are not normally distributed, but follow a skewed distribution bounded from below at $\Omega_{x,y} = 0$. To stabilize the variance we use $\log_{10}(1 + \Omega_{x,y})$ instead of just $\Omega_{x,y}$ in all our analyses. Furthermore, in our main analysis we restrict the area around the focus individuals to a circle with 10 m radius, which is a scale at which many ecologically important processes are expected to happen in rain forest trees and shrubs (e.g., Uriarte *et al.*, 2004). Even though different species might show interesting features in their pattern at different scales, we had no *a priori* reason to adapt the scale of the analysis for individual species. Moreover, Condit *et al.* (2000) showed that the Ω values at different scales are highly correlated.

2.2.2 Modelling dioecious species

We developed an IBM which we use to model how local abundance and sex-ratio in dioecious species might affect the aggregation pattern (similar to the model we use in Chapter 3

to explore the relation between aggregation and abundance and change in abundance). The landscape is assumed to be homogeneous, with spatial aggregation emerging only from dispersal processes of the modelled species, and not from habitat selection. We do not model any interactions between species or individual trees. The location of each individual tree is given by continuous x- and y-coordinates.

The model takes the sex ratio s and a local abundance n as parameters and then runs a simulated birth and death process. The model is initialized by randomly placing n individuals within the simulated area following a uniform distribution for both the x- and the y-coordinate. Each individual is randomly assigned to be male or female with a probability depending on the sex ratio s. After a given number of time steps the resulting aggregation pattern is analysed by calculating the Ω aggregation index. In each time step, one individual (male or female) is selected at random to die, and one of the female individuals is selected to reproduce. The reproducing female gives rise to an offspring at a position that is determined by a negative exponential radial dispersal kernel with a mean dispersal distance of 10 m and a dispersal angle from a uniform distribution between 0° and 360°. The sex of the offspring is chosen randomly depending on the sex ratio s. The results are not strongly dependent on the particular shape of the tail of the dispersal kernel as we show in Chapter 3. Though the total population size is kept constant throughout the simulation, the ratio of males and females is allowed to fluctuate. However, because the sex ratio of new offspring is fixed, while the likelihood of males or females to be selected for removal from the population varies with the relative abundance of the two sexes in the population, the realised sex ratio has a trend to return to the sex ratio parameter s. Setting the sex ratio parameter s to one (i.e., 100% females) allows us to simulate non-dioecious species in which all individuals are potentially able to produce seeds.

To ensure a constant local abundance and to avoid edge effects, we used periodic boundary conditions with individuals being dispersed out of the simulated area entering the area on the opposite side. We define one generation as n time steps, the expected number of birth/death events it takes until an individual is replaced by a new one. The simulation is run for 200 generations which is sufficient for the model to converge towards its natural equilibrium and to lose its dependence on the initial distribution of the individuals (see Figure 2.1).

2.2.3 Data collection in tropical rainforests

2.2.3.1 The Center for Tropical Forest Science (CTFS) plots

The Center for Tropical Forest Science (CTFS, 2013) is an initiative by the Smithsonian Tropical Research Institute that coordinates research forest dynamic plots around the world in cooperation with local partners. 48 forest plots are part of the CTFS network, mainly in the tropical regions, with most between 25 and 50 ha in size. The research on the different plots is conducted by different universities and research institutes, but they all use the same methods to facilitate across site comparisons (Condit, 1998). In each forest, all trees and shrubs with

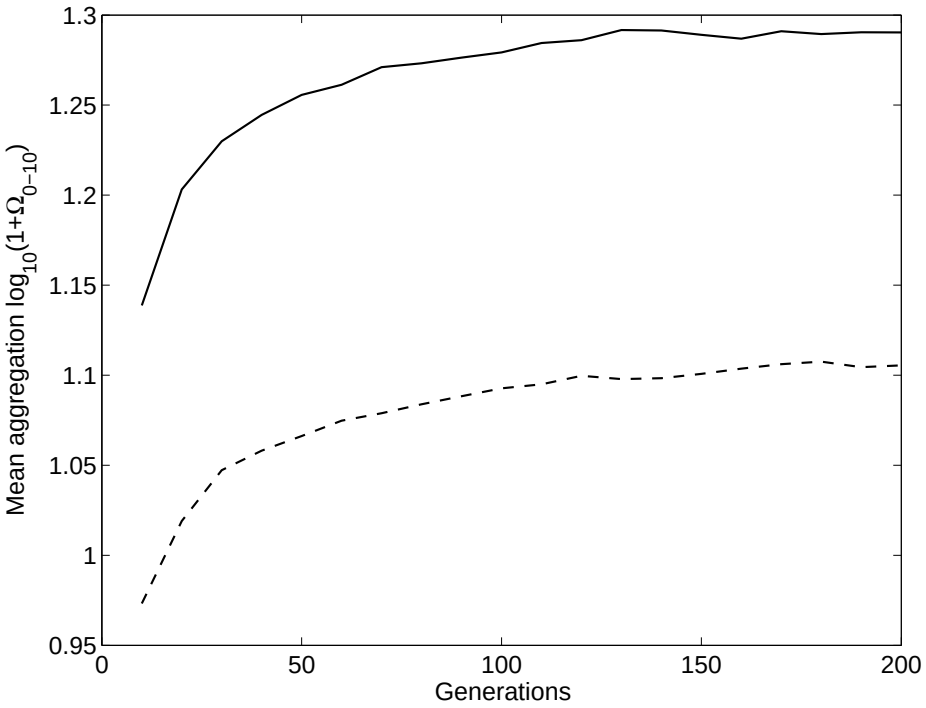


Figure 2.1: The graph shows how the simulation loses the initial dependence on the (uniformly random) start distribution and converges to its natural equilibrium. The solid line shows the mean $\Omega_{0,10}$ of 1000 runs of the simulation with an abundance of $n = 500$ and a sex ratio of $s = 0.5$. The dashed line show the results for a non-dioecious species ($s = 1$) with the same abundance.

a stem diameter at breast height (DBH; with breast height defined as 1.3 m or 50 cm above any buttresses, whatever is higher) of at least 1 cm are mapped, their DBH is measured and they are identified to species level. Regular re-censuses of the same forest should enable the investigation of forest dynamics and the CTFS aims to re-census each forest plot every five years. When a forest plot is established the area is divided into 20-by-20 m quadrants, and each corner of the quadrant is marked by a pole marked with an east/west and an north/south coordinate to allow orientation in the plot (see Figure 2.2). The topography of the plot is established by determining the altitude at all quadrant corners. Each 20-by-20 m quadrant is subsequently further subdivided into 5-by-5 m sub-plots, for which the corners are marked by unmarked smaller poles. After the plot is sub-divided, each tree and shrub with a stem diameter of more than 1 cm DBH is tagged with a number plate that assigns a unique id to each individual plant, the DBH is measured and its species identity is determined. The map of trees and shrubs may then form the basis for research in which further information is collected such as concentrations of soil minerals, seeds in seed traps or the growth of seedlings below 1 cm DBH.

The CTFS forest network is unique in its geographic spread of the forest plots around the world and the area size of the fully mapped plots. With more than 100,000 individuals in individual plots, this data is particularly well suited to study the processes within an individual

Figure 2.2: Photo of the Center for Tropical Forest Science (CTFS) 50 ha forest dynamic plot at Yasuní National Park, Ecuador. In the center of the photo a pole marking the corner of a 20x20 m quadrant is shown.

forest site. Other research initiatives such as rainfor in South America (rainfor, 2013) and AfriTRON in Africa (AfriTRON, 2013) have established large numbers of forest plots with typically only 1 ha size, which are more suitable to compare forest structure and species composition over larger geographical distances. The Smithsonian Institution Global Earth Observatory (SIGEO) works on expanding the CTFS network to include temperate forests and to study carbon fluxes and investigate the effects of climate change on the forests.

2.2.3.2 Yasuní

The forest dynamic plot in Yasuní National Park, Ecuador, was established in 1995. Initially, the censused area comprised 25 ha but was later expanded to 50 ha. Re-censuses have been completed in 2002 and 2007. The principal investigator of the plot is Dr. Renato Valencia of the Pontificia Universidad Catolica del Ecuador. Geographically it is situated in lowland Amazonian rainforest in eastern Ecuador in one of the most diverse regions of the globe with more than a thousand woody plant species within the 50 ha plot. Multiple new tree species have for the first time been described during the establishment of the Yasuní forest dynamic plot. The research plot is within a protected area, but oil extraction is done nearby and there is some hunting by indigenous people who live in the area. Most of the forest is old growth forest, but a small part of the plot had been cleared for a helicopter landing pad in the past. Within the forest dynamic plot an ongoing long term seedling study is being conducted, where seed fall is measured and seedling growth is monitored. Previous work by Simon Queenborough suggests that about 16-28% of species are dioecious. He studied all 16 species

of the dioecious family of Myristicaceae (nutmeg) in the forest plot over multiple flowering periods to determine the sex of the individuals and their reproductive investments. He found that female individuals made a much larger investment in reproduction than male individuals, but he did not find any common compensatory mechanism for that higher investment by females that was shared by all species. Some species showed male biased sex-ratios, in some species males started producing flowers at smaller sizes and in some species, male individuals would produce flowers more frequently. Female trees showed habitat correlations more frequently than males, but again this was not observed in all species and there was no systematic spatial segregation between sexes.

During a six week field trip in summer 2011, I studied the seven species of *Cecropia* found in the 50 ha forest dynamic plot at Yasuní (*C. sciadophylla*, *C. ficifolia*, *C. herthae*, *C. engleriana*, *C. putumayonis*, *C. marginalis*, and *C. membranacea*). *Cecropia* are a family of dioecious fast-growing pioneer species. I surveyed each *Cecropia* tree from the 2007 census, measured its DBH, and determined if it showed male or female flowers (see Figure 2.3). I also rated the health of each individual and assigned them to one of five classes; 0 for dead or disappeared trees, 1 for severely damaged trees with broken tops and/or which were completely covered by lianas without any leaves left, 2 for damaged individuals with broken branches, covered by lianas or which had lost many/most leaves, 3 for individuals with minor damage that were tilted, had one liana growing on them, or had lost some leaves, and 4 for healthy individuals with no recorded damage. Because individuals below 10 cm DBH at the 2007 census only very rarely showed flowers, I concentrated my efforts on the 586 individuals with a DBH larger than 10 cm. In addition I examined 145 *Cecropia* trees outside the mapped forest plot along the road to the research station, to compare sex ratios and the DBH of flowering trees in the closed canopy forest with trees in the high light conditions along the road. For some trees it was not possible to measure the DBH, generally either because liana coverage was too strong or because their buttresses where too high for me to reach the height at which the measurement was supposed to be taken according to the field protocol. For the individuals for which I could not measure DBH I estimated the change in DBH by fitting a linear regression model. I used the data of the trees of the same species and same sex for which DBH measures could be obtained and fitted a model with the DBH measured by me as dependent variable and the DBH determined at the 2007 census as explanatory variable (one outlier for which the DBH measured in 2011 was much smaller than the DBH in 2007 was excluded for the fit). I then used the parameters from that model to predict the DBH of the unmeasured trees based on their 2007 DBH measurements. I also computed the within-species as well as the within males and within females co-association pattern $\Omega_{0,10}$ of the most common *Cecropia* species in the plot.

Figure 2.3: The left image shows male *Cecropia* flowers, the right image female flowers. Male flowers are characterized by many small spikes, whereas females consist of fewer but larger spikes.

2.2.3.3 Barro Colorado Island (BCI)

The 50 ha forest dynamic plot at BCI, Panama, was the first forest plot established using the CTFS methodology. It was fully censused for the first time in 1982-1983. Starting in 1985 it has been re-censused every five years. Barro Colorado Island was cut off from the mainland by the construction of the Panama Canal. It is part of a protected nature reserve and hunting is banned from the island since the 1980s. Most of the forest plot is old growth forest, although about 2 ha had been cleared for farming about 100 years ago. About 300 different species of trees and shrubs can be found within the plot.

In his book on the flora of Barro Colorado Island, Croat (1978) estimated that about 9% of plant species on the island are dioecious. Of the 143 species we used in our analysis (tree and shrub species with a mean abundance of at least 100 individuals above 1 cm DBH in the 50 ha plot; see also Chapter 3 for more details) 35 are classified as dioecious by him.

2.2.4 Statistical analyses

We compute the $\Omega_{0,10}$ measure for all tree and shrub species in our analysis of the BCI data and then use two linear regression models to explore the amount of variance in the aggregation that is explained by the abundance of a species and by whether a species is known to be dioecious or not. The error terms in the linear regression models $\boldsymbol{\epsilon_1}$ and $\boldsymbol{\epsilon_2}$ are vectors which are assumed to have zero mean and are independent between species. The first model assesses how much of the variation in the log aggregation indices $\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}})$ of the species at BCI can be explained by the logarithm of the local abundance $\log_{10}(\boldsymbol{N})$ (the logarithm is here always defined element-wise on the vector of values; $\mathbf{1}$ is a vector of ones):

$$\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}}) = \beta_0 \mathbf{1} + \beta_1 \log_{10}(\boldsymbol{N}) + \boldsymbol{\epsilon_1}. \tag{2.2}$$

The second linear regression model assesses how much of the variation in the log aggregation indices can be explained by both $\log_{10}(\boldsymbol{N})$ and the binary variable $\boldsymbol{\zeta}$ indicating if a species is dioecious:

$$\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}}) = \beta_0 \mathbf{1} + \beta_1 \log_{10}(\boldsymbol{N}) + \beta_2 \boldsymbol{\zeta} + \boldsymbol{\epsilon_2}. \tag{2.3}$$

Clearly (2.2) is a special case of (2.3) with $\beta_2 = 0$. In both models $\boldsymbol{\Omega_{0,10}}$, $\boldsymbol{N}$ and $\boldsymbol{\zeta}$ are vectors with one entry per species included in the analysis. We assumed that the variance of the residual error would be dependent on the local abundance and therefore used weighted least squares to fit the parameters of the model (Sheather, 2008, p. 115). We used the weighting factors that we estimated as described in Chapter 3.

For each model we compute the amount of variation in $\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}})$ explained by the explanatory variables (R^2). To test whether the linear relationship between an explanatory variable and the response variable explains a significant proportion of the total variation, we used bootstrapping (Wassermann, 2006) to re-sample the data under the assumption of no relationship. Using bootstrapping is necessary to estimate significance intervals because we cannot assume normality of the error terms. By re-sampling from the data we can make inferences on the significance of the effects found in a forest with the same marginal distribution of the variables as BCI; the significance of the same results in a different forest could differ.

In the bootstrapping procedure for the first model (Equation 2.2), 10,000 random samples are created by independently and randomly drawing $\Omega_{0,10}$ and local abundance values N from the empirical distribution given by the data. Comparing the percentage of variation explained in the bootstrap samples under the assumption of independence and the percentage of variation explained in the true model allows the null hypothesis of independence of $\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}})$ and $\log_{10}(\boldsymbol{N})$ to be rejected if the explained variation in the true model is among the highest 5% of the re-sampled results (significant), or among the highest 1% (highly significant). We found 10,000 random samples were sufficient to obtain stable results.

In the second model we test the null hypothesis that there is a linear relationship between the binary variable $\boldsymbol{\zeta}$ indicating if a species is dioecious and $\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}})$ given the abundance $\boldsymbol{N}$. The same bootstrapping method described above is used, this time drawing pairs of $\Omega_{0,10}$ and local abundance values from the data, and then independently drawing $\boldsymbol{\zeta}$ values. This tests whether it is possible to reject the null hypothesis, and thereby establishes whether there is a linear relationship between $\boldsymbol{\zeta}$ and $\log_{10}(\mathbf{1} + \boldsymbol{\Omega_{0,10}})$ which is not already accounted for by $\log_{10}(\boldsymbol{N})$.

2.3 Results

2.3.1 The effect of breeding system in the individual-based model

Our model predicts that dioecious species are more aggregated compared to non-dioecious species of the same abundance and with identical seed dispersal kernel. Figure 2.4 shows the mean aggregation of 1000 runs of the simulation comparing a dioecious species with a sex-

ratio of 50% female and a non-dioecious species in which all individuals are able to produce seeds. In both cases aggregation is negatively correlated with abundance, with results for low abundances showing a higher variance.

Figure 2.5 shows how the aggregation depends on the ratio between male and female individuals in the population. We again compare the aggregation of dioecious species with non-dioecious species, but rather than focusing on populations with equal total abundance, we compare populations with equal expected numbers of seed producing individuals. In the non-dioecious species this number is equal to the total abundance, but in the dioecious species it depends on the ratio of males to females in the population. In the case of the dioecious species we can also look at the aggregation of the female and the male sub-population individually. The results of the simulation show that the mean aggregation of the female sub-population in a dioecious species is equal to the mean aggregation in a non-dioecious species with the same number of individuals. However, the aggregation of only the males or the whole population is lower because the male individuals do not form nuclei for clusters of new off-spring and are therefore more isolated.

Taken together the results from Figure 2.4 and Figure 2.5 imply that we should expect dioecious species to be more aggregated than non-dioecious species of the same abundance, but less aggregated than non-dioecious species with the same number of seed producing individuals.

2.3.2 Cecropia species in Yasuní, Ecuador

Of the 586 *Cecropia* trees above 10 cm DBH in the 2007 census of the forest plot, in 2011, 222 were found to be dead or heavily damaged. Even though *Cecropia* are a fast growing pioneer species, this was an unexpectedly high loss of individuals within a period of only 4 years, which might be explained by heavy storms in the time between the 2007 census and my field work. Only very few individuals of the second most common species *C. ficifolia* were found to be flowering. This might be due to the population suffering from wide-spread parasites, because many *C. ficifolia* had damaged leaves. But it could also be related to the fact that smaller mid-canopy species such as *C. ficifolia* may suffer from insufficient light exposure in such a closed canopy forest environment. Another possibility could be that it was not the main flowering season for *C. ficifolia*. Because of the large number of dead trees and because not many *C. ficifolia* were found flowering, only for *C. sciadophylla* could sufficient data be collected to enable any form of statistical analysis. The number of individuals of different species flowering in the different health classes is recorded in Table 2.1.

The first notable result is that for *C. sciadophylla* the number of male and female flowering individuals in the forest plot was nearly even. The data for other species is too sparse to draw any strong conclusions. However, I note that 5 out of 7 flowering *C. ficifolia* within the forest plot and 3 out of 4 along the road were female, whereas for *C. herthae* 8 out of 9 flowering individuals in the plot and both individuals found along the road were male (see

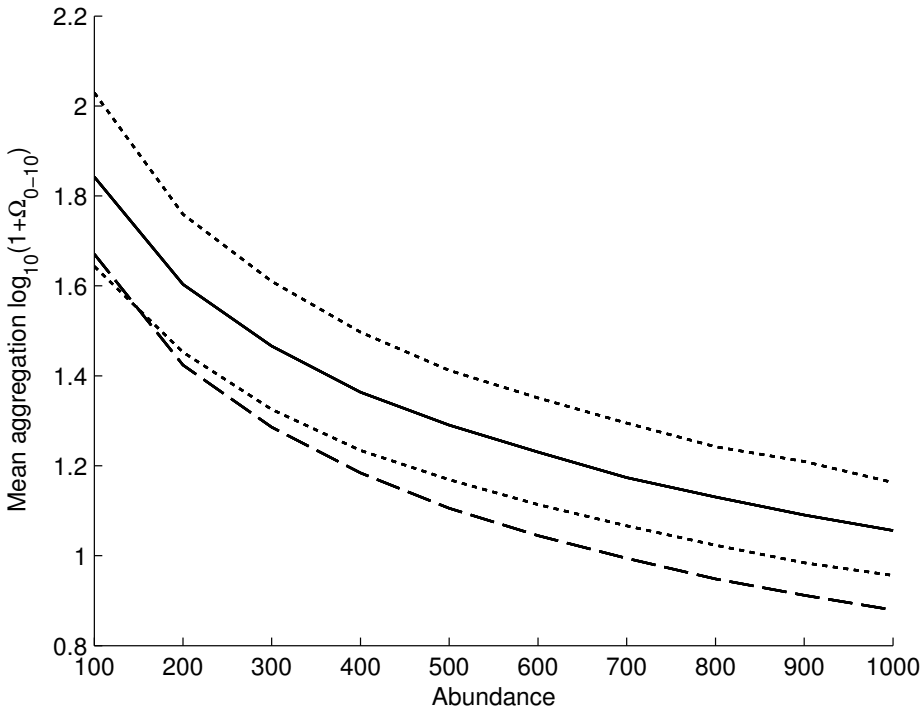


Figure 2.4: The graph shows the relationship between population abundance and aggregation for simulated dioecious species (solid line) with equal chance for individuals to be male or female, and non-dioecious species (dashed line). The dotted lines show the 5 and 95 percentile of the simulation results for the dioecious species. All results are showing the mean $\Omega_{0,10}$ at the end of 1000 runs of the simulation after generation 200.

Species code	Number of trees with that health level (top total, left female, right male)											
	0		1		2		3		4		$\sum$	
C. sciadophylla	134		4		18		16		171		343	
	0	0	0	0	2	1	2	6	52	51	56	58
C. ficifolia	46		10		10		6		80		152	
	0	0	0	0	0	0	0	0	5	2	5	2
C. herthae	9		0		3		1		24		37	
	0	0	0	0	0	0	0	0	1	8	1	8
C. engleriana	9		0		1		2		22		34	
	0	0	0	0	0	0	2	0	5	8	7	8
C. putumayonis	3		1		0		0		5		9	
	0	0	0	0	0	0	0	0	0	0	0	0
C. marginalis	2		1		0		0		1		4	
	0	0	0	0	0	0	0	0	0	0	0	0
C. membranacea	3		0		0		0		1		4	
	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.1: *Cecropia* surveyed within the 50 ha forest dynamic plot in Yasuní National Park, Ecuador. Health levels were defined as: 0 for dead/disappeared; 1 for severely damaged [broken and/or completely covered in lianas without leaves left]; 2 for damaged individuals [broken, liana covered or which had lost many/most leaves]; 3 for individuals with minor damage [tilted, with one liana, or which had lost some leaves], and 4 for individuals for which no damage was recorded.

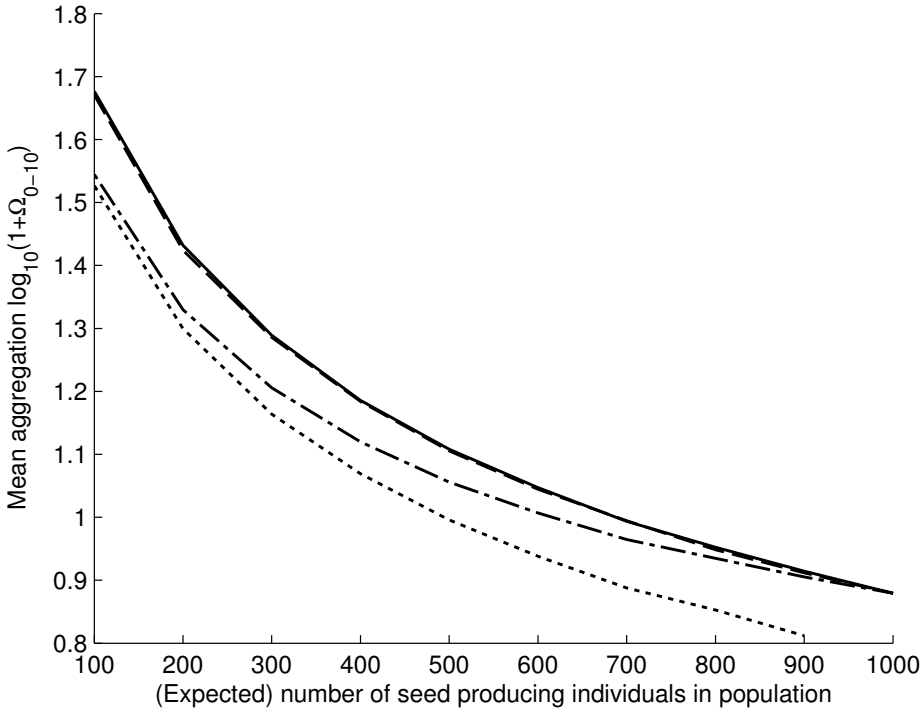


Figure 2.5: The graph shows the relationship between sex ratio and aggregation for simulated non-dioecious species with varying abundance (solid black line), and for dioecious species with a total abundance of 1000 individuals and varying sex-ratio (black dotted line show the aggregation for only the male individuals of the dioecious species, the dashed line for only the females, and the dashed-dotted line for all individuals). All results are showing the mean $\Omega_{0,10}$ at the end of 1000 runs of the simulation after generation 200.

Species code	Number of trees with that health level (top total, left female, right male)											
	0		1		2		3		4		$\sum$	
C. sciadophylla	0		3		15		23		77		118	
	0	0	0	0	2	3	3	4	16	35	21	42
C. ficifolia	0		0		0		2		7		9	
	0	0	0	0	0	0	1	0	2	1	3	1
C. herthae	0		0		1		1		5		7	
	0	0	0	0	0	0	0	0	0	2	0	2
C. engleriana	0		0		0		2		4		6	
	0	0	0	0	0	0	1	1	3	1	4	1
CECR??	0		1		0		0		4		5	
	0	0	0	0	0	0	0	0	0	2	0	2

Table 2.2: *Cecropia* surveyed along the road. Health levels were defined as as shown in Table 2.1. The line marked CECR?? reports the results for all individuals with undetermined species identity.

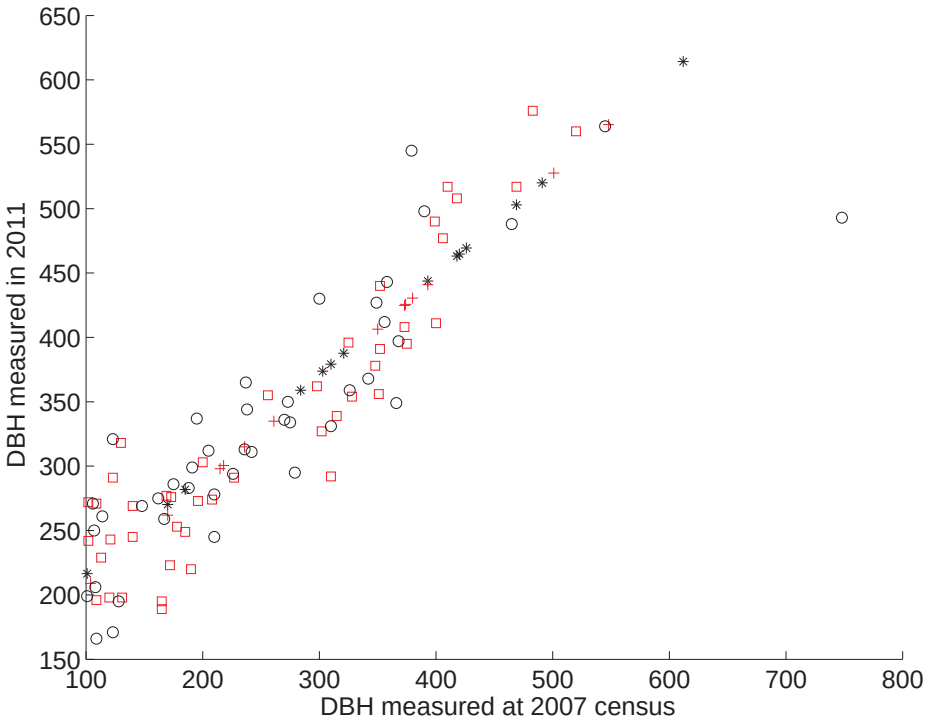


Figure 2.6: The Figure shows the DBH in 2007 compared to the DBH in 2011 for all flowering *C. sciadophylla* trees within the 50 ha forest dynamic plot as Yasuní, Ecuador. Female individuals for which DBH was measured in 2011 are represented by a black O, males with a □. For individuals that could not be measured in 2011, the DBH value for 2011 was estimated by a linear fit to the data of the individuals that could be measured (one outlier for which the DBH measured in 2011 was much smaller than the DBH in 2007 was excluded for the fit). Females with estimated 2011 DBH values are depicted by a black $*$, males with a $+$.

Table 2.2 for the results of the road census). This suggest that sex ratios and flowering frequency of the two sexes might differ between the species. Along the road there were twice as many flowering male compared to female *C. sciadophylla* trees (42 males compared to 21 females). This suggests that the flowering frequency of different sexes might be differentially affected by light availability. There is no large difference with respect to the DBH of the flowering individuals between male and female *C. sciadophylla* (see Figure 2.6). The mean DBH of flowering females within the plot was 34.9 cm (s.d. 10.3), for males it was 34.1 cm (s.d. 10.7). The mean DBH of the flowering *C. sciadophylla* along the road was slightly lower (33.0 cm for females and 33.7 cm for males), but there was no indication that the larger proportion of males flowering could be explained by an earlier onset in flowering for the males. As expected from the simulation model we found males to be slightly less aggregated than females. The $\Omega_{0,10}$ relative neighbourhood co-association measure was 7.8 within all individuals, but 10.2 within the female sub-population and only 6.6 within the male sub-population.

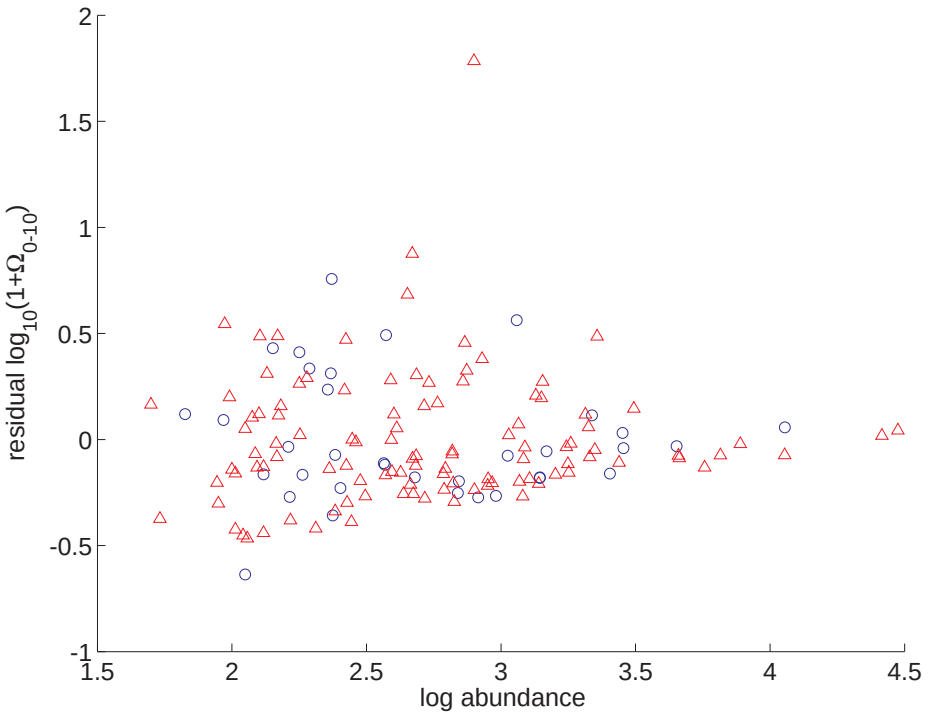


Figure 2.7: The graph shows the residual aggregation of the linear regression model after accounting for current abundance. Non-dioecious species are represented by Δ; dioecious species by O. There does not seem to be a systematic difference in aggregation related to breeding system.

2.3.3 The effect of breeding system at Barro Colorado Island

Abundance is a highly significant predictor for within-species spatial aggregation. The first linear regression model with only a constant factor and abundance as explanatory variable explains 13.5% of the variance (R^2) in the $\log_{10}(1 + \Omega_{0,10})$ data. This is more than the variance explained in any of the 10,000 randomised bootstrap samples (see Chapter 3 for further analyses of the relation between abundance and spatial aggregation).

Adding the categorical variable on whether a species is dioecious as explanatory variable, does not contribute to explaining further variance in the aggregation. The explained variance of the second linear regression model which includes abundance, the variable on whether a species is dioecious, and a constant factor is slightly higher with a R^2 value of 13.6%. However, this improvement is less than what would be expected from adding a random second variable. The result of the bootstrap randomization shows that 58.5% of the 10,000 random samples have a higher explained variance than the model containing the actual information on breeding system. Figure 2.7 shows the residual $\log_{10}(1 + \Omega_{0,10})$ of the species after subtraction of the fit of the first linear regression model with respect to abundance and breeding system of species. There does not seem to be a systematic relation between aggregation and whether or not a species is dioecious.

2.4 Discussion

It has been suggested based on previous models that dioecy should lead to higher within-species aggregation if dioecy is not compensated by other mechanisms (Bleher *et al.*, 2002; Heilbuth *et al.*, 2001). There is empirical evidence that this expected difference in aggregation is not found in a smaller forest plot in Costa Rica (Hubbell, 1979). However previous studies have not directly compared simulation results with empirical data. The results of our IBM support the hypothesis that dioecy should, all else being equal, increase the aggregation of a species. We achieve this with a very simple mechanistic model that still produces realistic patterns similar to those found in nature (see also Chapter 3). Our IBM results are directly comparable to data from the 50 ha CTFS forest dynamic plots. The effect is of a magnitude that we would expect to detect an increased mean aggregation among dioecious species. However, we find no difference in within-species aggregation between dioecious and co-sexual species. Importantly, we do not even see a small effect which could have suggested that we are merely lacking statistical power or that the expected pattern is masked by other processes affecting the community. There is no evidence for any influence of dioecy on aggregation. This clearly suggests that dioecious species must have evolved traits that ensure less aggregation, for example superior means of seed dispersal, rather than superior competitiveness under high intra-specific competition. This conclusion is supported by Queenborough *et al.* (2009) who found that dioecious species have no advantage in seed mass or abundance in Yasuní, which would have suggested advantages of dioecious species in survival or seed production. However, our results cannot show which process or trait leads to the smaller than expected aggregation, nor can they show that this is the only compensatory advantage of dioecious species. There are other potential mechanisms besides superior seed dispersal that could lead to the observed patterns of aggregation. For example, dioecious species could have higher tolerance to certain environmental conditions allowing them to be habitat generalists compared to non-dioecious species. Alternatively, species could start reproduction earlier in life, or a larger percentage of individuals might reach reproductive age, the proportion of seed producing individuals in that case might not be as different from co-sexual species as would be predicted solely based on the sex-ratio.

Queenborough *et al.* (2007) has shown that different species use different mechanisms to adapt to the particular challenges of dioecy. Some show equal sex ratios, while others are skewed towards male individuals. In some species females show different or stronger habitat associations than males. In the field study we conducted on *Cecropia* at Yasuní we also found differences between sex ratios between different *Cecropia* species, although this result should be treated with care, given the small number of individuals among all species except *C. sciadophylla*. On a more stable numerical basis, we did however observe a very different sex ratio between *C. sciadophylla* within the closed canopy forest and the high light condition along the road, with a much higher percentage of males along the road. This is

a surprising result as females generally have higher costs in reproduction, so we might have expected relatively more females to flower in the high light condition. This result can also not be explained by an earlier onset of flowering in males compared to females along the road. *C. sciadophylla* flower at slightly smaller DBH along the road than they do in the forest, but we did not observe a difference in the size of the flowering individuals between males and females. We did however observe that females of *C. sciadophylla* were more aggregated than males, as predicted by our IBM. *Cecropia* species are also very well-dispersed canopy trees which supports our tentative conclusion from the IBM and the BCI data that dioecious species rely more on good dispersal than on coping with stronger intra-specific competition. The fact that *Cecropia* are fast growing pioneer species, which generally have a higher proportion of adult individuals in the population (Wright *et al.*, 2003), might also help to explain how they might compensate for the fact that only part of the adult population produces seeds. To determine for a larger number of species which traits enable overcoming the disadvantages of dioecy will require more work on individual species as well as statistical analysis across many species. Our result point to a few traits, such as dispersal ability, percentage of adult population, and habitat width, that we consider likely to contribute to the persistence of dioecy in plants. To determine their relative importance remains an open task for future work.

Low species density may pose difficulties to dioecious species; it can for example lead to limited pollen being available if female trees are out of reach for male pollen. However, the degree to which dioecious species are disadvantaged due to higher conspecific aggregation may also depend on the overall species density. If the overall species density is low, conspecific competition due to increased aggregation might matter less. Interestingly, only about 6% of all flowering plants are dioecious, but about 9% of species at the moderately diverse tropical forest on Barro Colorado Island are dioecious, and about 16-28% of species at the bio-diversity hot-spot in the Yasuní forest are dioecious. This observation, that the percentage of dioecious species is higher where overall species densities are lower, suggests that conspecific competition due to the reduced number of potential seed dispersers may be the main evolutionary disadvantage dioecious species have to overcome. The prevalence of dioecy might be used as one of several indicators to determine the dominant evolutionary forces acting on species in future comparative studies of ecosystems.

Bibliography

AfriTRON. 2013. *African Tropical Rainforest Observation Network*. [online] http://www.geog.leeds.ac.uk/projects/afritron/. Accessed: 2013-06-15.

Bleher, Bärbel, Oberrath, Reik, & Böhning-Gaese, Katrin. 2002. Seed dispersal, breeding system, tree density and the spatial pattern of trees - a simulation approach. *Basic and Applied Ecology*, **3**(2), 115–123.

Condit, Richard. 1998. *Tropical Forest Census Plots*. Berlin, Germany, and Georgetown, Texas: Springer-Verlag and R. G. Landes Company.

Condit, Richard, Ashton, Peter S., Baker, Patrick, Bunyavejchewin, Sarayudh, Gunatilleke, Savithri, Gunatilleke, Nimal, Hubbell, Stephen P., Foster, Robin B., Itoh, Akira, LaFrankie, James V., Lee, Hua Seng, Losos, Elizabeth, Manokaran, N., Sukumar, R., & Yamakura, Takuo. 2000. Spatial Patterns in the Distribution of Tropical Tree Species. *Science*, **288**(5470), 1414–1418.

Croat, Thomas B. 1978. *Flora of Barro Colorado Island*. Stanford University Press.

CTFS. 2013. *Center for Tropical Forest Science*. [online] http://www.ctfs.si.edu. Accessed: 2013-06-15.

Heilbuth, Jana C., Ilves, Katriina L., & Otto, Sarah P. 2001. The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. *Evolution*, **55**(5), 880–888.

Hubbell, Stephen P. 1979. Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest. *Science*, **203**(4387), 1299–1309.

Hubbell, Stephen P., Condit, Richard, & Foster, Robin B. 2005. *Barro Colorado Forest Census Plot Data*. [online] https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Illian, Janine, Penttinen, Antti, Stoyan, Helga, & Stoyan, Dietrich. 2008. *Statistical analysis and modelling of spatial point patterns*. West Sussex, England: John Wiley.

Janzen, Daniel H. 1970. Herbivores and the Number of Tree Species in Tropical Forests. *The American Naturalist*, **104**(940), 501–528.

Queenborough, Simon A., Burslem, David F. R. P., Garwood, Nancy C., & Valencia, Renato. 2007. Determinants of biased sex ratios and inter-sex costs of reproduction in dioecious tropical forest trees. *Am. J. Bot.*, **94**(1), 67–78.

Queenborough, Simon A., Mazer, Susan J., Vamosi, Steven M., Garwood, Nancy C., Valencia, Renato, & Freckleton, Rob P. 2009. Seed mass, abundance and breeding system among tropical forest species: do dioecious species exhibit compensatory reproduction or abundances? *Journal of Ecology*, **97**(3), 555–566.

rainfor. 2013. *Amazon Forest Inventory Network*. [online] http://www.rainfor.org/. Accessed: 2013-06-15.

Renner, Susanne S, & Ricklefs, Robert E. 1995. Dioecy and its correlates in the flowering plants. *American Journal of Botany*, 596–606.

Sheather, Simon J. 2008. *A Modern Approach to Regression with R*. Springer Texts in Statistics. Dordrecht: Springer.

Uriarte, M., Condit, Richard, Canham, C. D., & Hubbell, Stephen P. 2004. A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? *Journal of Ecology*, **92**, 348–360.

Wassermann, Larry. 2006. *All of nonparametric statistics*. New York: Springer.

Wright, S Joseph, Muller-Landau, Helene C, Condit, Richard, & Hubbell, Stephen P. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology*, **84**(12), 3174–3185.

Chapter 3

The relationship of abundance and within species aggregation

Summary

The current spatial pattern of a population is the result of previous individual birth, death and dispersal events. We present a simple model followed by a comparative analysis for a species rich plant community to show how the current spatial aggregation of a population may hold information about recent population dynamics. Previous research has shown how locally restricted seed dispersal often leads to stronger aggregation in less abundant populations than it does in more abundant populations. In contrast, little is known about how changes in the local abundance of a species may affect the spatial distribution of individuals. If the level of aggregation within a species depends to some extent on the abundance of the species, then changes in abundance should lead to subsequent changes in aggregation. However, an overall change of spatial pattern relies on many individual birth and death events and a surplus of deaths or births may have short-term effects on aggregation that are opposite to the long-term change predicted by the change in abundance. The change in aggregation may therefore lag behind the change in abundance, and consequently the current aggregation may hold information about recent population dynamics. Using an individual-based simulation model with local dispersal and density dependent competition, we show that on average, recently growing populations should be more aggregated than shrinking populations of the same current local abundance. We test this hypothesis using spatial data on individuals from a long-term tropical rain forest plot, and find support for this relationship in canopy trees but not in understory and shrub species. On this basis we argue that current spatial aggregation is an important characteristic that contains information on recent changes in local abundance, and may be applied to taxonomic groups where dispersal is limited and within-species aggregation is observed.

Impact of this work

The work presented in this chapter is one of few studies providing an example on how individual-based computational models can inform ecological theory. We show that analysing static patterns may not only shed light on static properties of a system, such as rates of competition between species and the influence of environmental factors on species distribution, but may also reveal dynamic changes the ecosystem is undergoing. Being able to determine the dynamic changes ecosystems experience, while lacking data on the past state of those systems, might prove essential in the protection of our natural heritage in a changing world.

Declaration on the contributions to the work presented in this chapter:

This chapter is published as:

The memory of spatial patterns: changes in local abundance and aggregation in a tropical forest, Anton J. Flügge, Sofia C. Olhede, and David J. Murrell, Ecology, vol. 93, nr. 7, p. 1540–1549, 2012.

The computational model presented in this chapter was developed by me. All analysis were conducted by me. The first draft was written by me. Sofia Olhede and David Murrell contributed by supervising and guiding my work and by revising the manuscript before publication. The data used was provided by the Centre for Tropical Forest Science.

3.1 Introduction

The spatial pattern of individuals within a local population is the result of the many processes that determine the birth, death and movement (dispersal) rates of those individuals (see Dale, 1999; Levine & Murrell, 2003). This means that the current spatial distribution of a population could hold information on its demographic history. However, because there are many processes that could lead to the same pattern, and because multiple biological processes are likely to be operating at the same time, linking pattern and process has proved to be a difficult task (Levin, 1992). Nonetheless, linking spatial pattern and demography is an important goal in ecology because it offers an opportunity to investigate likely recent population dynamics when time series of abundances are not available.

The spatial aggregation of species within a local population is an important characteristic and is directly related to many ecological processes, such as dispersal (Carlon & Olson, 1993; Seidler & Plotkin, 2006), competition (Kenkel, 1988; He & Duncan, 2000), and habitat selection (Getzin *et al.*, 2008). Much effort has gone into describing the spatial patterns of populations and communities. Plant communities, among them rain forests, have

been intensely studied (e.g. Condit *et al.*, 2000). One of the common results is that, even in forests without strong habitat inhomogeneities, populations with higher local abundance are less aggregated than populations with lower local abundance (Condit *et al.*, 2000; Wang *et al.*, 2010). This is true, even though more recently it has been shown that rare species at Barro Colorado Island (BCI) may be suffering stronger conspecific competition (Comita *et al.*, 2010) and, according to the Janzen-Connell hypothesis (Janzen, 1970), conspecific competition should give rise to a less aggregated spatial pattern. Computer simulations have suggested spatially localized dispersal may be sufficient to cause rare species to be more aggregated than abundant populations (Bleher *et al.*, 2002); and more formal spatial extensions to the logistic equation have shown how the scales of competition and dispersal interact to generate the level of aggregation or spatial segregation (Bolker & Pacala, 1997; Law *et al.*, 2003).

The observation that rarer species are on average more aggregated than locally abundant species suggests that there might be a link between the spatial pattern and past population dynamics. In particular, this relationship suggests that a species that has been declining in abundance should have been less aggregated in the past (when it was more abundant), and a currently equally abundant species that has been increasing in abundance should have been more aggregated in the recent past (when it was rarer). Spatial ecological theory has made explicit the relationships between individual births, deaths, dispersal and emergent spatial pattern (Bolker & Pacala, 1997; Law *et al.*, 2003). Using spatial extensions to the familiar logistic equation for population growth, Law *et al.* (2003) have shown that short-term increases in aggregation are associated with births, particularly if dispersal is short-range, whereas decreases in aggregation are associated with deaths of individuals that have crowded neighborhoods (see equation 5 of Law *et al.* 2003). However, in the long-term, if rarer species are indeed more aggregated, then a surplus of deaths would need to lead to more aggregation and a surplus of births to less aggregation. The conflicting short- and long-term effects of change of abundance could also be modeled as a Thomas point process model (Illian *et al.*, 2008), in which individuals are distributed around a number of cluster centers, or parent individuals. In such a model, the death of individuals within a cluster reduces aggregation and the addition of individuals to clusters increases aggregation. In contrast, the establishment of new clusters decreases clustering as the landscape becomes more packed, and the extinction of existing clusters increases the population-level aggregation. If the establishment and death of individuals occurs on a shorter time scale than the establishment and death of clusters, the long-term trend in the aggregation of a population could be masked or even reversed in the short-term. Any time-lag between change in abundance, and long-term change in aggregation would mean that the expected current pattern of growing and declining populations would be different, with the growing population more aggregated than the declining population. If such relationships could be found, this would provide important information about which

species have been recently declining, and exploring these relationships clearly has practical implications for species conservation.

The concept that past changes in abundance could affect the current aggregation of a population has rarely been explored further, either theoretically or empirically. In rare exceptions, Wilson *et al.* (2004) have shown for butterflies, and Pocock *et al.* (2006) for plants, that the distribution of rare species throughout Great Britain correlates with past changes in their population size; with shrinking populations being more scattered than stable or growing populations. Contrary to Condit *et al.* (2000) and Wang *et al.* (2010), Wilson *et al.* (2004) and Pocock *et al.* (2006) found that more abundant species were more aggregated. This difference might be due to the different spatial scales in the two sets of studies since Wilson *et al.*'s (2004) and Pocock *et al.*'s (2006) data was on occupancy of 10 km squares across an entire country, whereas Condit *et al.* (2000) and Wang *et al.* (2010) consider point pattern data for individuals and consider aggregation over tens of meters. The difference could, however, also be due to the different aggregation indices used in the studies, as the relationship between abundance and aggregation will be sensitive to how aggregation is defined and measured: both the intrinsic summary and its statistical properties will be important. These differences aside, Wilson *et al.* (2004) and Pocock *et al.* (2006) did find a relationship between current aggregation and recent demography at the large spatial scale, but to our knowledge, the link between changes in abundance and aggregation have yet to be investigated at the scale of individuals.

In this study we explore how past local abundance and recent population growth or decline might leave its mark on the pattern of individuals within a population. We begin with an individual-based simulation model to provide a mechanistic explanation of how a change in aggregation can lag behind a change in local abundance, and in particular how declining populations might be less aggregated than recently growing populations of the same current abundance. Based on these theoretical results we then look for evidence of such relationships at the local scale in the data from the BCI long-term forest dynamic plot (Condit, 1998; Hubbell *et al.*, 1999, 2005), and find empirical support for the memory of recent changes in local abundance in the current spatial distribution of individuals.

3.2 Materials and methods

3.2.1 Measure of aggregation

We are using the same $\log_{10}(1 + \Omega_{0,10})$ relative neighbourhood density measure that was introduced in Section 2.2.1. By using the same aggregation measure as Condit *et al.* (2000), our results are more easily comparable to their analysis. We test the robustness of our main results by varying the scale of Ω and report how that affects our results.

3.2.2 Modelling the effect of abundance and change of abundance

We use an individual-based model (IBM) to investigate how dispersal distance and local abundance affect the aggregation pattern of a species. The landscape is assumed to be homogeneous, which means that heterogeneity in the abiotic landscape is not considered and the spatial aggregation emerges only from births and deaths of individuals, and not from habitat selection. Furthermore, we simulate each species on its own, thereby assuming heterospecific interactions to be insignificant. The location of each individual tree is given by continuous x- and y-coordinates, and this means the landscape is represented as a continuous field, rather than a discrete set of sites on a lattice.

The model takes a mean dispersal distance d and a local abundance n as parameters and then runs a simulated birth and death process. The model is initialized by randomly placing n individuals within the simulated area following a uniform distribution for both the x- and the y-coordinate, and after a given number of time steps the resulting aggregation pattern is analyzed by calculating the Ω aggregation index. In each time step, two individuals are selected at random, of whom one is selected to die and the other one is allowed to reproduce. The reproducing individual gives rise to an offspring at a position that is determined by a negative exponential radial dispersal kernel with a mean dispersal distance d and a dispersal angle from a uniform distribution between 0° and 360°. The results are not strongly dependent on the particular shape of the tail of the dispersal kernel, and in additional results we show the robustness of the results to a fat-tailed dispersal kernel. This birth process models the dispersal, emergence and establishment of a new individual, and it is assumed that new individuals are able to reproduce immediately at birth. This represents the simplest spatial birth-death process with local dispersal and a fixed population size. In contrast to models in which population size is allowed to drift (Felsenstein, 1975), the expected spatial aggregation will eventually stabilize in our model.

To ensure a constant local abundance and to avoid edge effects, we used periodic boundary conditions with individuals being dispersed out of the simulated area entering the arena on the opposite side. We define one generation as n time steps, the expected number of birth/death events it takes until an individual is replaced by a new one. The simulation is run for 200 generations which is sufficient time for the model to converge towards its natural equilibrium and to lose its dependence on the initial distribution of the individuals (see Figure 2.1). To simulate a population growth or decline, additional individuals are selected in each generation to be culled or allowed to reproduce.

3.2.2.1 Locally density dependent mortality

Just as dispersal is localized in space, ecological interactions tend to be strongest between individuals that are nearby in space. As an extension of the basic IBM we therefore also developed a locally density dependent (LDD) IBM. In the LDD IBM, we include an additional term, that ensures the selection of the individual that dies is dependent on the local density

of the population. This term adjusts the probability of an individual to be chosen to die by a factor that sums the contribution of nearby neighbors to the death rate of the focal individual. The effect of neighbors on the death rate decays exponentially with distance in the model (as in the statistical model of Comita *et al.*, 2010, supplementary material), meaning individuals that are far away from the focal individual have a much lower effect on its death rate than neighbors that are nearby. An additional parameter, f, of the model allows us to adjust the strength of the density dependence with small values of f signifying strong density dependence and large values of f signifying weaker density dependence.

The procedure to compute the likelihood for a tree to be chosen to die in the LDD IBM is as follows:

1. The Euclidean distance between all pairs of trees is computed. (This step is computationally expensive which makes it difficult to run the extended model for larger populations.)

2. Using an exponential kernel, the distances are transformed into weights which reflect how much the trees influence one another, i.e. trees closer to one another have a stronger effect. The shape of this kernel is of the form $e^{-0.2 \cdot \text{distance}}$ (Comita *et al.*, 2010, report this function as describing the relationship between distance and density dependent mortality).

3. The weights are summed up for all trees (i.e. the sum of the weights between tree x and all other trees is computed for all trees x in the population), yielding a measure of how much competition $d0_x$ each tree x has to endure.

4. The values are then adjusted by subtracting the smallest weight from all weights such that the smallest value is now zero. We call those values d_x.

5. Another parameter of the model, f, specifies how strongly the density should influence which tree is chosen. For each tree x we compute a measure $l_x = d_x + f * \text{mean}(d)$ (i.e. for $f = 0$, l_x will be identical to d_x, whereas for larger values of f, the relative differences between the values l_x for the trees will be decreased and the density will have less of an effect).

6. Finally, a tree is chosen at random with the likelihood being dependent on the values of l for all trees (i.e. if l_x is twice as big as l_y, tree x is chosen twice as likely as tree y).

The procedure can be adjusted to simulate density dependent reproduction and allow facilitation (positive density dependence) instead of competition (negative density dependence).

3.2.3 Data used from Barro Colorado Island (BCI)

The 50 ha forest plot at BCI was established in the early 1980s and is one of the best studied forests in the world (see Condit, 1998; Hubbell *et al.*, 1999, 2005). A complete census was

first conducted in 1982-1983, and has since then been repeated every five years from 1985 onwards. In each census the position and species identity of all trees and shrubs with more than 1 cm diameter at breast height (DBH) was determined and more than 200,000 individual stems and 298 different species were identified in the 2005 census. In our analysis we only consider species with an average of over 100 individuals between the 1985 and 2005 censuses (for a complete list of the species see Appendix 7.1; we excluded *Bactris major* as the methodology of counting individuals of *Bactris* species changed during the study period and abundance data from the early censi is not comparable, see Feeley *et al.* 2011). This is because the sampling variation in the aggregation index $\Omega_{0,10}$ becomes very large for populations with lower local abundance, and is likely to obscure any ecological signal that may be present. We used the data of all individuals above 1 cm DBH. Although aggregation might differ between different size classes (Murrell, 2009), this is not expected to bias the results since it has been shown that the size-structure of populations is only weakly correlated with changes in local abundance (Condit *et al.*, 1998; Feeley *et al.*, 2007). Species at BCI are classified as shrubs, understory, mid-canopy, or top-canopy trees depending on whether they are a shrub or tree species, and on the maximum size they can reach, and we use this classification to group species in the statistical analyses. Since the BCI plot is protected from human intervention, most of the plot is old growth forest (Condit *et al.*, 1999) and it is considered a fairly homogeneous forest environment (Harms *et al.*, 2001), which suggests habitat dependence should only play a limited role in the aggregation of species. Taken together, the relatively long history of available data and its weak habitat inhomogeneities make this forest a particularly good test case for our theory that local dispersal and changes in local abundance induce detectable changes in the aggregation of a species. To test whether our results might still be explained by habitat inhomogeneities, we conducted an additional, more conservative test in which we only used those species for which Harms *et al.* (2001) could not show any habitat dependency at BCI.

3.2.4 Statistical analyses

We use two linear regression models. The first model assesses how much of the variation in the log aggregation indices $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ of the species at BCI can be explained by the logarithm of the current local abundance $\log_{10}(\boldsymbol{N_t})$ at time t:

$$\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}}) = \beta_0 + \beta_1 \cdot \log_{10}(\boldsymbol{N_t}) + \boldsymbol{\epsilon_{t,1}}. \tag{3.1}$$

The second linear regression model assesses how much of the variation in the log aggregation indices can be explained by both $\log_{10}(\boldsymbol{N_t})$ and the log of the relative changes in local abundance $\boldsymbol{m_{t,\Delta t}}$ (with $\boldsymbol{m_{t,\Delta t}} = \frac{\boldsymbol{n_t}}{\boldsymbol{N_{t-\Delta t}}}$) between time $t - \Delta t$ and t:

$$\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}}) = \beta_0 + \beta_1 \cdot \log_{10}(\boldsymbol{N_t}) + \beta_2 \cdot \log_{10}(\boldsymbol{m_{t,\Delta t}}) + \boldsymbol{\epsilon_{t,2}}. \tag{3.2}$$

Clearly (3.1) is a special case of (3.2) with $\beta_2 = 0$. In both models $\boldsymbol{\Omega_{0,10,t}}$, $\boldsymbol{N_t}$ and $\boldsymbol{N_{t-\Delta t}}$ are vectors with one entry per species included in the analysis. $\boldsymbol{\epsilon_{t,1}}$ and $\boldsymbol{\epsilon_{t,2}}$ are vectors of error terms with zero mean which are assumed independent between species. We assumed that the variance of the residual error would be dependent on the current local abundance and therefore used weighted least squares to fit the parameters of the model (Sheather, 2008, p. 115). We estimated the weighting factors by fitting a power function to the residual errors of the unweighted linear regression using the current local abundance $\log_{10}(\boldsymbol{N_t})$ and relative abundance change between 2000 and 2005 as main effects (as explained below).

For each model we compute the amount of variation in $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ explained by the explanatory variables (R^2). To test whether the linear relationship between an explanatory variable and the response variable explains a significant proportion of the total variation, we used bootstrapping (Wassermann, 2006) to re-sample the data under the assumption of no relationship. Using bootstrapping is necessary to estimate significance intervals because we cannot assume normality of the error term. By re-sampling from the data we can make inferences on the significance of the effects found in a forest with the same marginal distribution of the variables as BCI; the significance of the same results in a different forest could differ.

In the bootstrapping procedure for the first model (Equation 3.1), 10,000 random samples are created by independently and randomly drawing $\Omega_{0,10,t}$ and current local abundance values N_t from the empirical distribution given by the data. Comparing the percentage of variation explained in the bootstrap samples under the assumption of independence and the percentage of variation explained in the true model allows the null hypothesis of independence of $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ and $\log_{10}(\boldsymbol{N_t})$ to be rejected if the explained variation in the true model is among the highest 5% of the re-sampled results (significant), or among the highest 1% (highly significant). We found 10,000 random samples were sufficient to obtain stable results.

In the second model we test the null hypothesis that there is a linear relationship between the relative changes in local abundance $\log_{10}(\frac{\boldsymbol{n_t}}{\boldsymbol{N_{t-\Delta t}}})$ and $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ given the current abundance $\boldsymbol{N_t}$. The same bootstrapping method described above is used, this time drawing pairs of $\Omega_{0,10,t}$ and local abundance values from the data, and then independently drawing relative abundance change values. This tests whether it is possible to reject the null hypothesis, and thereby establishes whether there is a linear relationship between $\log_{10}(\frac{\boldsymbol{n_t}}{\boldsymbol{N_{t-\Delta t}}})$ and $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ which is not already accounted for by $\log_{10}(\boldsymbol{N_t})$. In the results below, we present analyses of the regression models for all 143 species that fulfilled our population size criteria, and separately for just the 94 top- and mid-canopy species among them.

3.2.4.1 Estimating weights for the linear regression model

The variance of the residuals in our linear regression model is related to the abundance of species, as $\Omega_{0,10}$ values for rare species show much more variance than the aggregation of

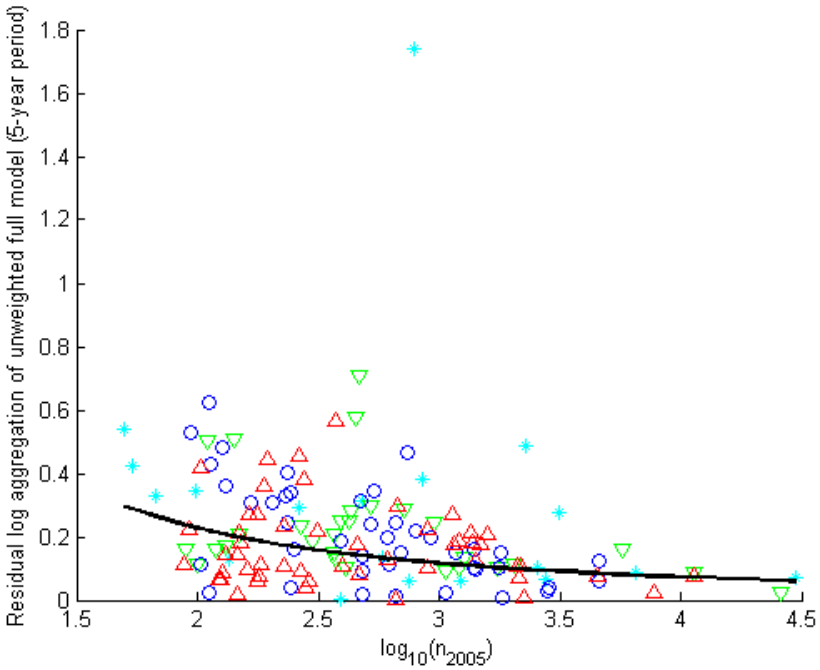


Figure 3.1: The relationship between local abundance and the residuals of the linear regression model for all 143 species from BCI that meet our selection criteria (top-canopy trees Δ; middle canopy trees O; understory trees ∇; shrubs $*$). The black line shows the estimated standard deviation $\sqrt{w_i}$ of the error that was used for the weighted regression in the our analysis.

more common species. We therefore used a power function to fit the variance of the residuals and use the fitted estimates as a weighting factor in the weighted linear regression. Figure 3.1 shows the residuals and the fitted function. We modeled the variance as:

$$\text{var}\left\{\epsilon_{2005,i}\right\} = a^2(\log n_{2005,i})^{2b}. \tag{3.3}$$

This implies that:

$$\log\|\epsilon_{2005,i}\| = \log a + b\log\|\log n_{2005,i}\| + \nu_{2005,i}, \tag{3.4}$$

where $\nu_{2005,i}$ has constant mean and variance. We fit a and b using least squares, and then use weighted regression with this fitted model, with a weighting function given by:

$$w_i = \frac{1}{\hat{a}^2(\log n_{2005,i})^{2\hat{b}}} \tag{3.5}$$

One outlier observation is ignored in this fit.

3.2.4.2 Checking for phylogenetic independence of species

Using the linear regression model we implicitly assume independence between species. We test whether we should correct for shared phylogenetic history between species by investi-

gating the potential for phylogenetic relatedness to explain part of the variation in the aggregation index of the BCI species we analyse. Kress *et al.* (2009) produced a phylogeny of the Barro Colorado Island woody plant species based on DNA data, and we use this phylogeny to compute phylogenetic linear models using the caper R-package version 0.4 (http://cran.r-project.org/web/packages/caper/index.html). To test whether the phylogeny provides information for this analysis we used caper to compute the maximum likelihood value of Pagel's λ (Pagel, 1997, 1999) for all four phylogenetic linear models in our analysis (change for the 5-year and the 20-year period, and using all species or only the canopy species). Pagel's λ is a measure of whether the structure of a phylogenetic tree can well explain the data under a constance-variance random walk model of evolution, with λ values of around one meaning that this model well explains the data, while a λ value close to 0 indicates that the data is better explained if species are assumed to have evolved independent of each other.

Not all species in our analysis were included in the phylogeny and some species had zero distance in the phylogeny. We replaced species with zero distance by their most recent common ancestor for which we assumed that it had the mean value of its successor species. Of the 143 species in our analysis we got to a phylogenetic tree with 126 tips (species or common ancestors of species with zero distance) and 80 tips for the phylogeny of only the canopy species.

The maximum likelihood estimates for Pagel's λ were zero for all models with the upper bound of the 95% confidence intervals between 0.18 and 0.28 (as computed by caper using likelihood ratio tests). As none of our models showed a phylogenetic signal we concluded it would be justified to assume independence of species and not correct for phylogeny in our main results.

3.3 Results

3.3.1 Simulation results

Our IBM shows a negative correlation between aggregation and local abundance, replicating the simulation results of Bleher *et al.* (2002). Interestingly, the model can produce an abundance-aggregation relationship as found in the empirical data when using a range of mean dispersal distances as has been reported by Muller-Landau & Hardesty (2005); see Figure 3.2. We note that in our simulation model the dispersal kernel describes the successful establishment of new individuals and so encapsulates both seed dispersal and seedling/sapling establishment, whereas the estimated range of the dispersal kernel concerns only the seed dispersal phase.

Analyses of the simulation model suggest that when local abundance changes there is a time lag before aggregation changes accordingly (Figure 3.3). The results clearly show that a growing or shrinking population is expected to have a different aggregation pattern than a stable population of the same current abundance; and in particular, growing populations

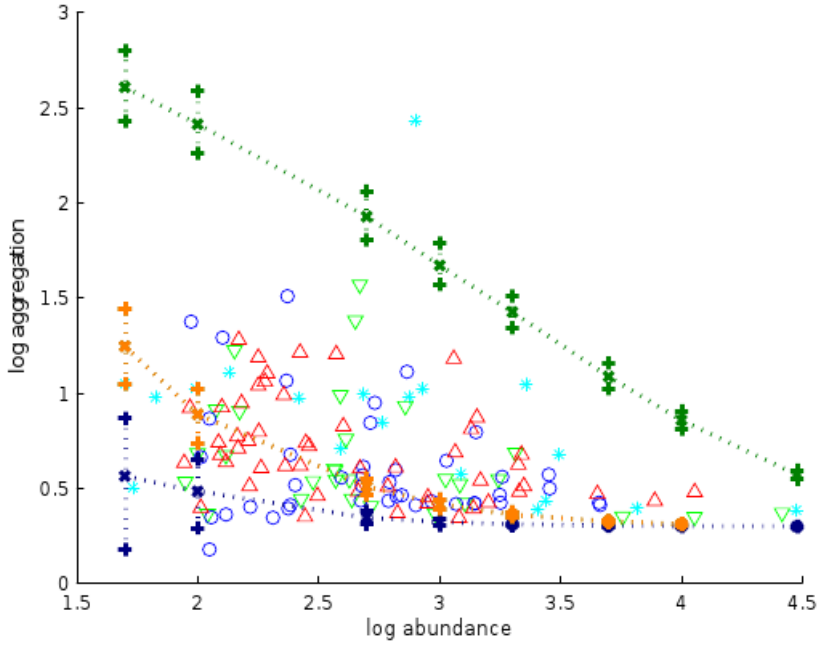


Figure 3.2: There is a strong correlation of abundance and aggregation of species (top-canopy trees Δ; middle canopy trees O; understory trees ∇; shrubs $*$). The green line shows the mean result of 500 runs of the individual-based model (for population sizes of 50, 100, 500, 1000, 2000, 5000, 10000, and 30000) with a mean dispersal distance of 2.8 m, the orange line the respective simulation results for a mean dispersal distance of 30 m, and the blue line for 152 m. The whiskers mark the 10 and 90 percentile of the simulation results. As reported by Muller-Landau & Hardesty (2005) species at BCI have mean dispersal distances between 2.8 m and 152 m, all but two species have $\Omega_{0,10,2005}$ values that simulations with a mean dispersal distance within that range can well explain.

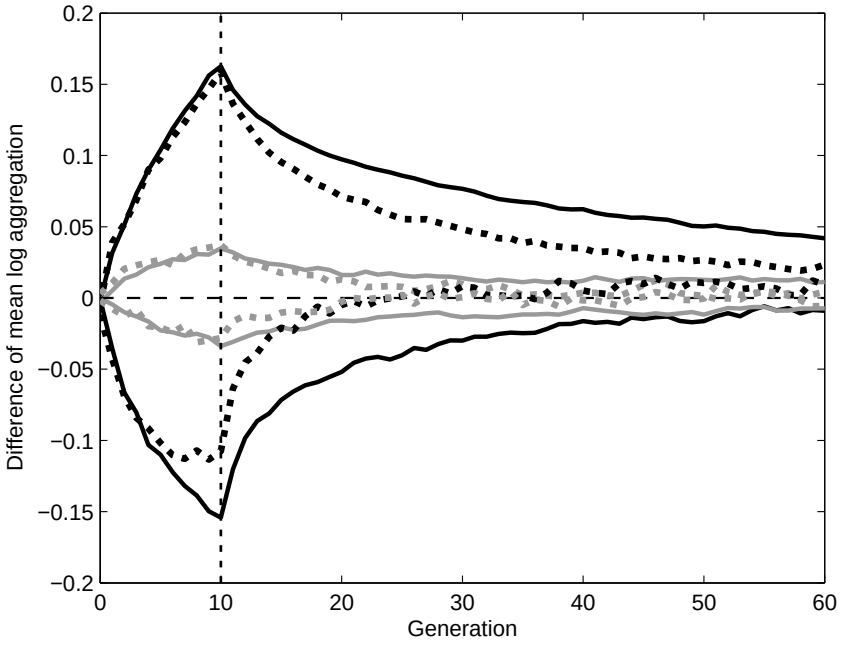


Figure 3.3: The graph shows the mean difference between the $\log_{10}(1 + \Omega_{0,10})$ value of a simulated population with changing local abundance, and a population with the same but stable abundance. Depicted is the mean of 1000 runs of the basic individual-based model (with a mean dispersal distance of 10 m). It was initialized with 100 (dotted) or 500 (solid) individuals in the population. Then, in the population with changing local abundance, the population was decreased (lines going below zero) or increased (lines going above zero) by 10% (black) or 2% (gray) per generation for a total of 10 generations. After 10 generations (vertical dotted line), the abundance was kept constant and one can observe how the mean $\log_{10}(1 + \Omega_{0,10})$ relaxes back to the expected value for a stable population of that abundance (horizontal dashed line).

are likely to be more aggregated and declining populations are likely to be less aggregated than stable populations. As expected the effect on the aggregation strongly depends on the magnitude of the abundance change, with small changes yielding small differences from a stable population; and in all cases, after the population stops growing/declining there is a lag phase before the aggregation converges on the $\Omega_{0,10,t}$ value for a stable population of the same abundance.

3.3.1.1 Locally density dependent mortality

Using the LDD IBM with a negative influence of local density on survival, we find that for stronger density dependence the model predicts a stronger effect of changes in local abundance on $\Omega_{0,10,t}$ in shrinking populations (see Figure 3.4). This is because local density dependence preferentially removes individuals that have crowded neighborhoods, leading to a rapid reduction in aggregation when deaths outweigh births; and therefore in the short-term, the difference between the aggregation of a decreasing population and a stable population is widened even further. However, in the extreme limits of strong locally density dependent mortality, the population may become spatially segregated to an extent that further mortality

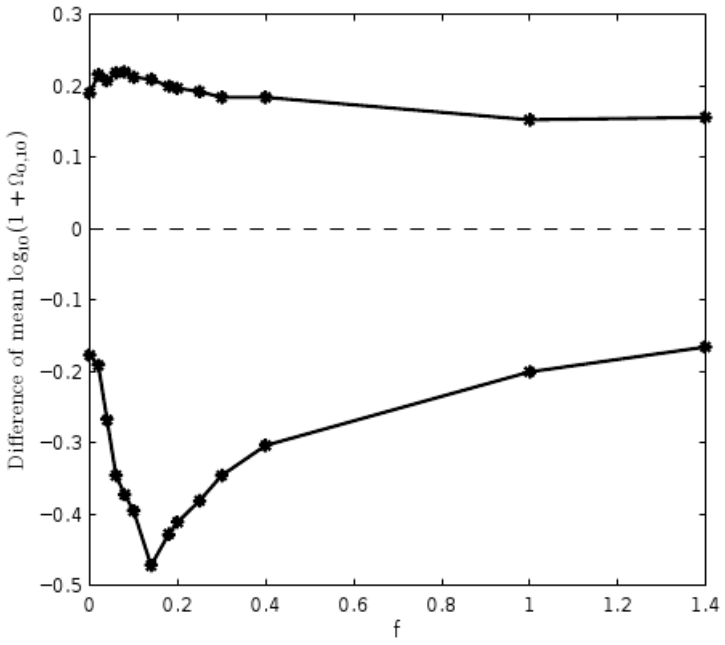


Figure 3.4: The graph shows the effect of changes in local abundance on aggregation, dependent on the strength of density dependent mortality f. Depicted is the mean difference between the $\log_{10}(1+\Omega_{0,10})$ value of a simulated population with changing local abundance, and a population with the same but stable local abundance after 10 generations. All simulations are based on at least 100 repetitions, with a starting population size of 100, a mean dispersal distance of 10 m, and 10% change in abundance during the first 10 generations. The upper line is based on the results of growing populations, the lower line on those of shrinking populations.

cannot lead to any further reduction of aggregation. In these cases the neighborhoods of most individuals are already empty with respect to conspecific individuals.

3.3.1.2 Effect of different dispersal kernel

In order to test the sensitivity of our model to the shape of the chosen dispersal kernel we also ran our simulation model using a Cauchy kernel, which has a location parameter that we set to 0, and a single scale parameter. The Cauchy kernel has a fatter tail than the negative exponential distribution, i.e. very long dispersal events are more likely. A consequence of the fatter tail is, for the same mean dispersal distance, the Cauchy distribution will produce more very short and more very long distance dispersal than an exponential dispersal kernel. We use a Cauchy kernel with a scale parameter of 2.5, but all other details of the individual-based model reported in the main text remain the same. This kernel has a slightly larger mean dispersal distance than the exponential kernel with a mean dispersal distance of 10 m that we use to generate the results in Figure 3.3, but still produces a more aggregated pattern, as even though long distance dispersal events are more likely, most dispersal events are over shorter distances.

The tail of a dispersal kernel can be important for the speed with which species can colo-

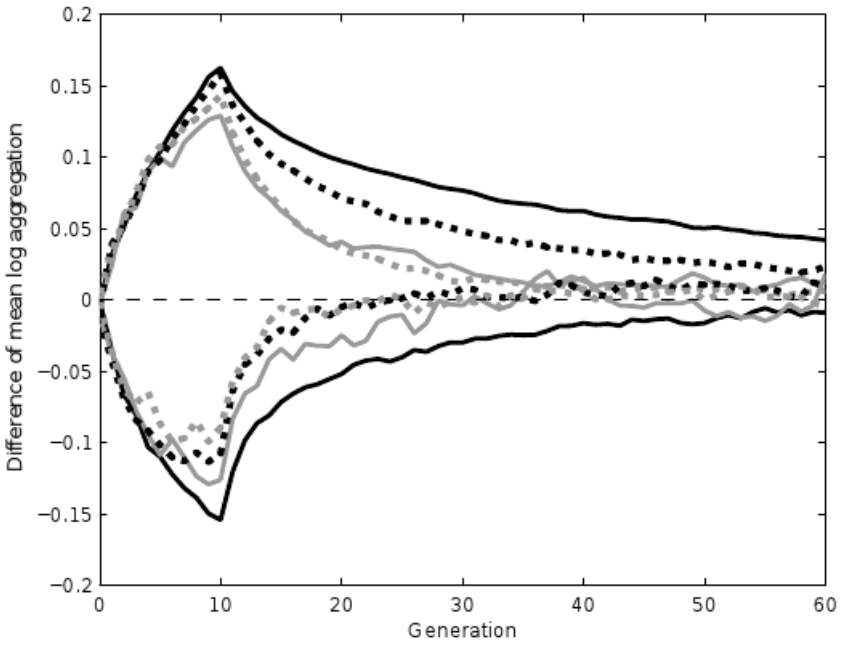


Figure 3.5: The graph shows the mean difference between the $\log_{10}(1+\Omega_{0,10})$ value of a simulated population with changing local abundance, and a population with the same but stable abundance. Depicted is the mean of 1000 runs of the basic individual-based model (100 runs for the Cauchy kernel with a starting population of 500). Each run was initialized with 100 (dotted) or 500 (solid) individuals in the population. Then, in the population with changing local abundance, the population was decreased (lines going below zero) or increased (lines going above zero) by 10% per generation for a total of 10 generations. After 10 generations, the abundance was kept constant and one can observe how the mean $\log_{10}(1 + \Omega_{0,10})$ relaxes back to the expected value for a stable population of that abundance (horizontal dashed line). Black lines show the results for the negative exponential kernel with a mean dispersal distance of 10 m, gray lines show the results for a Cauchy kernel with scale parameter of 2.5 (the expected aggregation of the simulation with the Cauchy kernel is higher than for a exponential kernel with 10 m mean dispersal distance, even though the mean dispersal distance is larger for the Cauchy kernel).

nize new areas, but our results show that it is not important in producing the local aggregation pattern at the scale at which we are analysing our data. Figure 3.5 compares the results for the aggregation of species with changing abundances in our simulation model using the Cauchy or the negative exponential kernel. There is no qualitative difference in our results for the two kernels.

3.3.2 Empirical results

There is a significant relationship between local abundance and aggregation in species at BCI, with rarer species on average being more aggregated (Figure 3.6). Fitting a weighted linear regression model to the log aggregation with the log abundance as main effect explains 17.9% of the variation (R^2) in the 143 species data (slope of the regression $\beta_1 = -0.20$), and 16.2% for only the 94 canopy species. This result is highly significant, with none of the 10,000 bootstrap samples showing a stronger correlation in the complete set and a p-

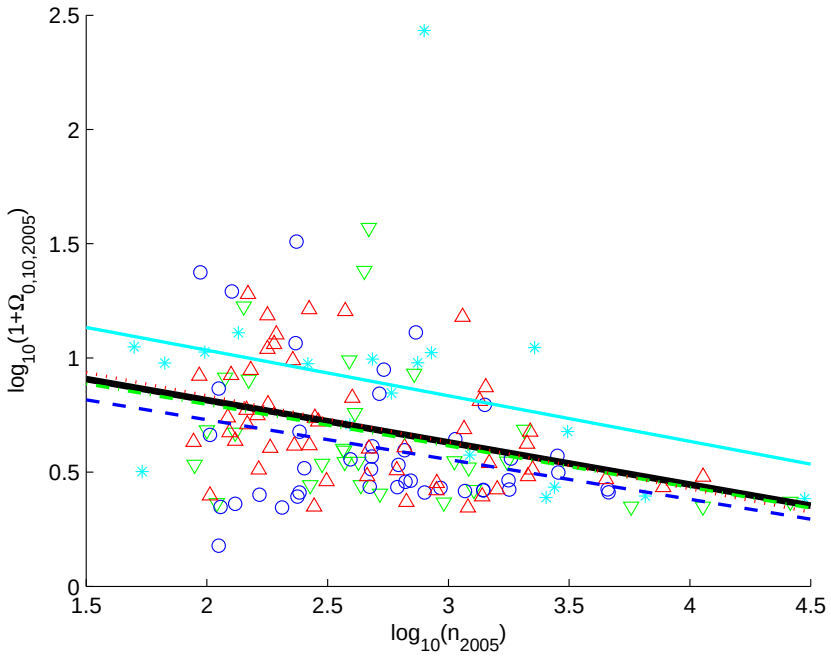
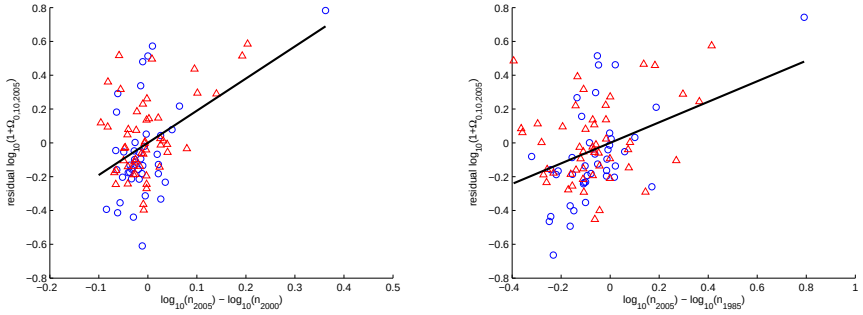


Figure 3.6: There is a strong correlation between local abundance $\log_{10}(\boldsymbol{N_{2005}})$ and aggregation of species $\log_{10}(1 + \boldsymbol{\Omega_{0,10,2005}})$ (top-canopy trees Δ; middle canopy trees O; understory trees ∇; shrubs $*$). The coloured lines show the best linear fit for the species of the four different growth types (solid light blue for shrubs, dashed-dotted green for understory trees, dashed dark blue for mid-canopy trees, and dotted red for top-canopy trees). The thick solid black line (behind dotted and dashed-dotted lines) shows the best linear fit of the data of all species.

value close to zero ($p < 0.001$) for the set restricted to only the canopy species. This finding confirms previous results that relate current abundance to spatial pattern (Condit *et al.*, 2000).

For two populations of the same current local abundance, but which have different recent population dynamics, the species which has been growing would necessarily have been rare compared to the species which has been declining in abundance. Figure 3.6 suggests that in the recent past, the locally growing (rare) species should have been more aggregated than the locally declining (common) species, and the results of our IBM suggest this past aggregation should persist even after some population change has taken place. We now turn to consider whether this difference in past aggregation has left a detectable signal in the current populations of the BCI community.

Fitting the linear regression model for all species using both the log abundance and the log relative abundance changes as main effects to predict the log aggregation indices $\log_{10}(1 + \boldsymbol{\Omega_{0,10,t}})$ (Equation 3.2) shows that adding the change in local abundance as an explanatory variable can explain an additional 10.0% of the variation in the aggregation indices (R^2) for the abundance changes between 2000 and 2005, and an additional 6.0% of the variation when considering the period between 1985 and 2005. Despite this, the results are not significantly better than the results obtained for the model using only the current abundance



(a) (b)

Figure 3.7: Plotted is the logarithm of the relative changes in local abundance (between 2000 and 2005 in (a) and 1985 and 2005 in (b)) against the residual $\log_{10}(1 + \mathbf{\Omega_{0,10,2005}})$ values after subtraction of the constant effect and the effect of current abundance. These were obtained from a weighted linear regression model with the logarithm of the local abundance and the logarithm of the relative changes in local abundance as main effects. The black line shows the remaining effect of the logarithm of the relative changes in local abundance. Top-canopy trees are depicted by Δ and mid-canopy trees by O. Note that (a) and (b) are depicted on different scales.

as main effect ($p = 0.076$ for the period 2000-2005; and $p = 0.21$ for the period 1985-2005).

We note that, given the same abundance, shrubs are on average more aggregated than canopy trees (Figure 3.6). This suggests that different processes may cause the aggregation in shrubs and canopy trees, and that grouping them together may mask these differences. Indeed, if we restrict our analyses to the 94 top- and mid-canopy tree species, we find that the model in Equation (3.2) can fit the data significantly better than the model in Equation (3.1). For only the canopy species, the model including local abundance changes can explain an additional 20.0% of the variation in $\log_{10}(1 + \mathbf{\Omega_{0,10,t}})$ ($p < 0.001$) for the 2000 to 2005 time span, and an additional 13.4% ($p = 0.014$) for the 1985 to 2005 time span. Figure 3.7 shows the relationship between the change in abundance and the residual aggregation of canopy species after subtraction of what could be explained by current abundances; it can be seen that species with increasing abundances on average are more aggregated than species with decreasing abundances even after correction for current abundance. This difference between the results for all species and that of only canopy species suggests there might be important biological differences between the canopy and the understory species, and we will discuss these in the following section. However, the results have to be interpreted with some caution because the mid-canopy tree species with the largest abundance change (*Cecropia obtusifolia*, a typical pioneer species, see Alvarez-Buylla & Martinez-Ramos 1992) has quite large leverage. Without this species, the model using the current local abundance and changes in the local abundance as main effects explains an additional 14.3% of the variation in aggregation ($p = 0.012$) for the time period 2000-2005, and an additional 8.2% of the variation ($p = 0.11$) for the period 1985-2000.

The range of changes in abundance was larger in the 20-year period, but the effect of

a change in abundance was smaller than in the 5-year model. For the longer period the β value corresponding to the changes in local abundance was 0.57, whereas the coefficient for the shorter period was 1.86. This shows that the memory of changes in abundance can be quite long for the top- and mid-canopy trees in the BCI dataset, but it also shows that the memory is diminished over longer time periods. This is presumably because in a longer time window there are more events that can affect the spatial pattern, and because the populations may undergo both positive and negative changes in abundance in the intervening years.

3.3.2.1 Non-habitat dependent species

To check if our results could be explained by habitat effects, we repeated our analysis for only those species which Harms *et al.* (2001) found to be not habitat dependent. Harms *et al.* (2001) tested 171 BCI species (those that had more than 65 individuals in the 1990 census) on whether they showed any habitat dependencies. Of these, 61 did not show any dependency with one of the 5 habitat types in the study (swamp, low plateau, high plateau, slope, stream-side), and 49 of these species are included in our study, 35 of which are top- or mid-canopy tree species.

There is no significant relationship between local abundance and aggregation in the 49 species identified by Harms *et al.* (2001) as not having any habitat dependency (see Figure 3.8). Fitting a weighted linear regression model to the log aggregation with the log abundance as main effect explains 5.6% of the variation in the 49 species data (R^2), and 7.2% for only the 35 canopy species. This result is not significant, with more than 17% of the 10,000 bootstrap samples showing a stronger correlation in the complete set (p = 0.17) and p = 0.15 for the set restricted to only the canopy species.

We fitted the weighted linear regression model for all species using both the log abundance and the log relative abundance changes as main effects to predict the log aggregation indices. This shows that the current local abundance and the changes in local abundance can explain 12.1% of the variation in the aggregation indices (R^2) for the abundance changes between 2000 and 2005, and 18.9% of the variation when considering the period between 1985 and 2005. Despite this, the results are not significantly better than the results obtained for the model using only the current abundance as main effect (p = 0.36 for the period 2000-2005; and p = 0.18 for the period 1985-2005). If we restrict our analyses to the 35 top- and mid-canopy tree species, we find that the model can fit the data much better. For only the canopy species, the model including local abundance changes can explain 32.1% of the variation in log aggregation (p = 0.061) for the 2000 to 2005 time span, and 32.4% (p = 0.055) for the 1985 to 2005 time span (see also Figure 3.9).

The variation explained for the non-habitat dependent canopy species is slightly less than for all canopy species when looking at the 5-year period, but slightly larger for the 20-year period. Because of the smaller sample set we have less statistical power, but in principle the results for the non-habitat dependent species seem to be in line with the results for all

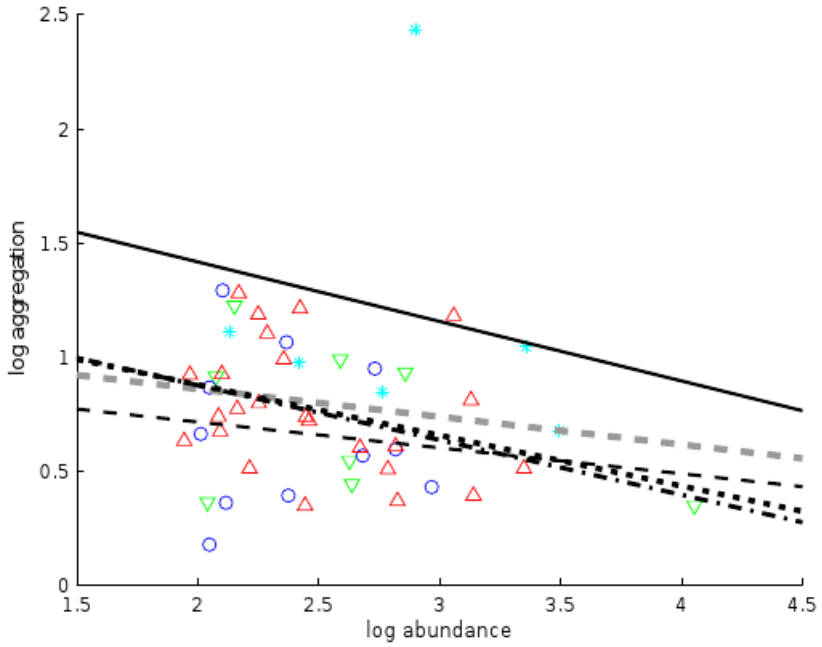
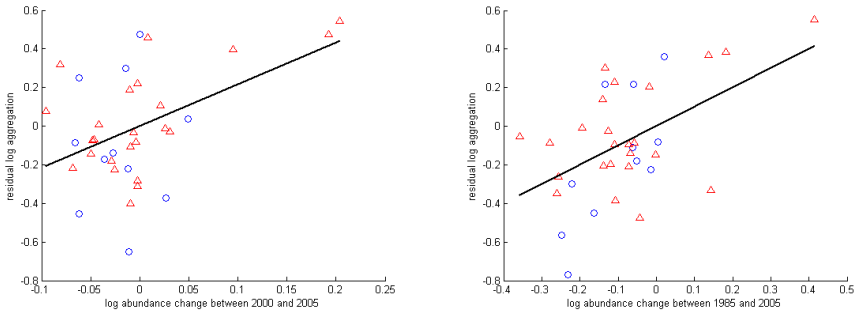


Figure 3.8: The relationship between local abundance and population aggregation for the 49 species from BCI that meet our selection criteria and that show no habitat dependency (top-canopy trees Δ; middle canopy trees O; understory trees ∇; shrubs $*$). The thick dashed gray line shows the best linear fit of the data of all species, the other lines show the best linear fit for the species of the four different growth types (solid for shrubs, dashed-dotted for understory trees, dashed for mid-canopy trees, and dotted for top-canopy trees). Using bootstrapping methods, we find the slopes are not statistically different from 0.



(a) (b)

Figure 3.9: Plotted is the logarithm of the relative changes in local abundance (between 2000 and 2005 in (a) and 1985 and 2005 in (b)) against the residual $\log_{10}(1 + \mathbf{\Omega_{0,10,2005}})$ values after subtraction of the constant effect and the effect of current abundance for those 35 canopy species that fulfilled our selection criteria and where not shown to be habitat dependent by Harms *et al.* (2001). These were obtained from a weighted linear regression model with the logarithm of the local abundance and the logarithm of the relative changes in local abundance as main effects. The black line shows the remaining effect of the logarithm of the relative changes in local abundance. Top-canopy trees are depicted by Δ and mid-canopy trees by O. Note that (a) and (b) are depicted on different scales.

species. That we fail to find the relationship between aggregation and abundance is also due to the fact that only very few of the most common species are non-habitat dependent, our sub-sample in this analysis is therefore biased towards rare species and represents a smaller range of abundances.

3.3.2.2 Effect of neighbourhood scale

We investigate the effects changing the scale of Ω has on the fits of the linear regression model with and without abundance change as explanatory variable. We repeat the analysis for the canopy species presented above, varying the outer radius y over which we measure $\Omega_{0,y}$; however, for simplicity, we do not weight the regression since earlier analysis shows this has only small effects on the results. As expected at Ω radii much larger than 10 m small scale aggregation patterns cannot be detected as well, but adding abundance change as explanatory variable can still significantly improve the explanatory power of the model (i.e. $p < 0.05$) for radii up to approximately 40 m. In Figure 3.10 we show the additional percentage of variance explained that is added by adding abundance change to the model together with the p-value for the significance test of whether the more complex model improves on the explanatory power. It should be noted that the values for Ω should be correlated with one another, since the information in Ω with small outer radii is also contained in the Ωs with larger outer radii. Results for radii smaller than 10 m are surprisingly good, which indicates that even though noise is expected to increase on average over all analysed species we can still detect a strong signal. The analyses over the 20 year period shows a similar pattern than the analysis over the 5 year period, however the effect of scale seems to be more pronounced over the shorter period.

3.4 Discussion

Ecological theory that incorporates spatially localized interactions and dispersal has implied that changes in abundance might be detected in the current pattern of a population, but this link has not been greatly explored theoretically or empirically (Bolker & Pacala, 1997; Law *et al.*, 2003). Analyses of our spatially explicit individual-based simulation model show that for populations of the same local abundance, and with identical dispersal kernels, a population that has recently been declining in abundance should be less aggregated than one that has been recently increasing in abundance (Figure 3.3). This pattern is created by two processes. Firstly, declining populations by definition were previously more common, and therefore were likely to be less aggregated than increasing populations that were relatively rare (see Figure 3.6). Any time-lag in the change in aggregation with the change in abundances will hence leave a memory of the initial aggregation in the current patterns (Figure 3.3). Secondly, deaths of individuals are often associated with strong neighborhood competition, and if mortality is caused predominantly by local conspecific competition, then this self-thinning could further reduce the aggregation of a declining population (see Figure 3.4).

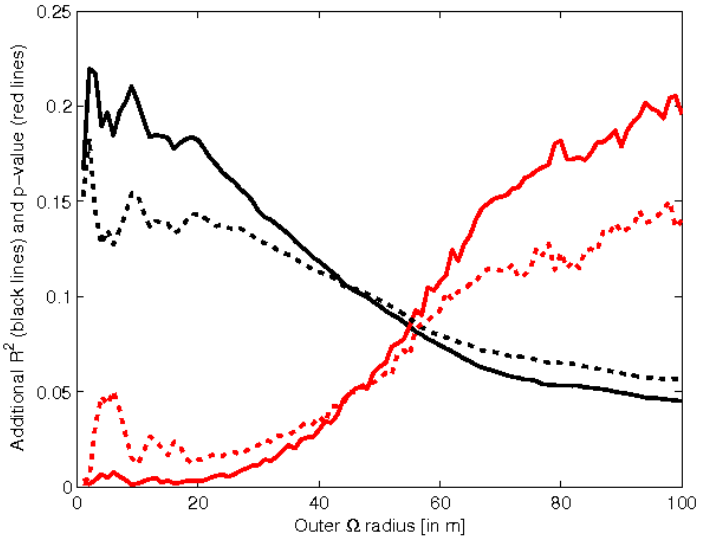


Figure 3.10: Black lines show the percentage of variance explained that is added by adding abundance change to the model (i.e. R^2 of the full model minus R^2 of the model with only the current abundance as explanatory variable). Red lines show the bootstrap p-value of whether adding abundance change to the model improves the model. Solid line show results for the 5 year period, dotted lines for the 20 year period. All results are based on the analysis of the canopy species.

Our results suggest that changes in abundance are important factors that should be considered when analyzing the spatial pattern of populations that proliferate pre-dominantly by local dispersal. Perhaps more intriguingly, it also suggests that the spatial pattern of a species could provide valuable information about the past history of a population. The regression model we built to investigate this relationship in the BCI tropical forest dataset shows some support for this hypothesis, but only when just the canopy trees are considered. Including all species in the same analysis reveals no significant relationship, and we argue this is due to differences in biology between canopy trees and shrubs.

Our simulation results also provide further support that the aggregation of species can be partly explained by the current local abundance, a fact that is well known (Hubbell, 1979; Condit *et al.*, 2000; Bleher *et al.*, 2002), but not always taken into account when analyzing the causes of aggregation. However, for canopy species the proportion of the variation in the aggregation index that can be explained by the relative changes in local abundance is about as large as the proportion of variation explained by the current abundance. This shows how the current spatial pattern of individuals may have a strong memory for recent population dynamics, something that has previously only been considered on larger spatial scales (Wilson *et al.*, 2004; Pocock *et al.*, 2006).

Analysis of the BCI dataset over the longest possible period shows that it is possible

to find a positive relationship between changes in local abundance over 20 years and current aggregation. However, the relationship is weaker because populations may have been fluctuating during this time frame, and therefore both increases and decreases in aggregation may have been occurring, and changes in abundance could have happened at any time during the 20 year period with short-term effects weakening in their influence the further back in the past the abundance change took place (Figure 3.7). The results of the simulation also suggest that persistent long-term trends should be detectable in the current spatial pattern when going back even further in time. The time scales at which we can expect to find effects of changes in local abundance depend on the rate at which individuals are replaced: the average generational turnover time. At BCI the average generational turnover of canopy species is only about 45 years (Hubbell *et al.*, 2005). In other forests, this time could be considerably longer, but in other communities such as grasslands (Seabloom *et al.*, 2005), a generation will be much shorter, and this should be taken into consideration when comparing communities with different dominant growth forms.

Given the many factors influencing aggregation patterns, and their high natural variability, at present we are not able to make definite claims on the history of a single species based on its aggregation pattern. However, when looking at species with very large increases of local abundance over the last five years, it is striking that all of them are more aggregated than expected for a species of their abundance (Figure 3.7). Therefore when investigating which species might be recent immigrants or which species are favored by recent environmental changes, it would be reasonable to start looking at those species which are more aggregated than expected. Finding reliable indicators of abundance change has proven to be difficult (see Condit *et al.* 1998 and Feeley *et al.* 2007 examining size distribution, an obvious candidate for predicting changes in local abundance), and it is likely that prediction may only be possible by considering several variables at once.

We found a significant positive relationship between changes in local abundance and aggregation in canopy trees but not in the set of all species. The difference between canopy and understory species might be caused by multiple processes. The ecological processes in the smaller growth-types might operate on different scales; they might suffer from different causes for mortality (Canham *et al.*, 2004); neighbor size might be more important than neighbor identity and smaller growth-types might therefore suffer more strongly from heterospecific competition and less from conspecific density dependence; and shrubs seem to be generally more aggregated (Figure 3.6) and less good seed dispersers than trees (Muller-Landau & Hardesty 2005 report that shrubs are more likely to be explosive dispersers). All these effects increase the noise in the data, and could interact with abundance and change of abundance in complex ways. Restricting the set of species to a more homogeneous subset of species eases those problems. Unfortunately, it is difficult to separately analyze the understory species, as there are only few shrub and understory tree species and we therefore have

very little statistical power.

We have considered only two main explanatory variables in our statistical model: current local abundance and recent changes in abundance. Among the other factors that could be considered when investigating the relationship of changes in local abundance and spatial pattern is habitat heterogeneity (Wiegand *et al.*, 2007; Getzin *et al.*, 2008), differences in size, age, and sex of individuals, life history strategies (Murrell, 2009), shade tolerance (Wang *et al.*, 2010), and drought tolerance of species (Feeley *et al.*, 2011), and dispersal syndrome (Muller-Landau & Hardesty, 2005; Seidler & Plotkin, 2006; Wright *et al.*, 2007).

Habitat heterogeneity is supposed to have little effect on most species at BCI (Harms *et al.*, 2001); thus, we considered it reasonable to use a model that does not include effects of habitat; but for a generalization of our results, habitat heterogeneity would be an important aspect in future analyses because it might interact with other explanatory variables (and we look at it in more detail in Chapter 4). In light dependent species new seedlings either grow very fast and reach adulthood quickly or die. In contrast, in shade tolerant species, individuals might be in a sapling stage growing only slowly and not reproducing for a long time. Therefore the percentage of reproducing individuals in a population will be lower in shade tolerant species than in light dependent species (Condit *et al.*, 1998) and consequentially shade tolerant species might be expected to be more aggregated. However, shade intolerant species might be more aggregated because their dependence on gaps means they might be aggregated at the sapling stage, and it remains to be seen if there is a general relationship between shade tolerance and spatial pattern. The breeding system of a species could also be an important variable in explaining variation in aggregation among species because dioecious species should have half the effective population size of an equally abundant hermaphrodite species, however we showed that there is no effect of breeding system on aggregation at BCI (see Chapter 2). Finally, the dispersal and establishment of seeds is a complex process which differs among species besides just different mean dispersal distances with for example animal dispersed seeds often being disposed in clumps of multiple seeds while wind dispersed seeds are distributed individually (Muller-Landau & Hardesty, 2005).

These caveats and extensions aside, we have provided theoretical and empirical support for the hypothesis that both current abundance, and recent changes in abundance leave their mark on the degree of aggregation in locally growing or declining populations. On this basis we believe it may indeed be possible to infer past population dynamics from current spatial patterns, and therefore link spatial pattern to process, a goal which has hitherto been hard to achieve in ecology (Murrell *et al.*, 2001). Since the described results are based on very basic mechanisms of local dispersal and neighborhood competition that are present in many ecosystems, this analysis could potentially be applied to a wide range of different systems, ranging from coral reefs (Karlson *et al.*, 2007) to the spread of diseases (Marshall, 1991). Future work should try to test whether the results are replicable on data from other ecosys-

tems, and could involve incorporating some of the suggested extensions where the additional information is available.

Bibliography

Alvarez-Buylla, Elena R., & Martinez-Ramos, Miguel. 1992. Demography and Allometry of Cecropia Obtusifolia, a Neotropical Pioneer Tree – An Evaluation of the Climax-Pioneer Paradigm for Tropical Rain Forests. *Journal of Ecology*, **80**(2), 275–290.

Bleher, Bärbel, Oberrath, Reik, & Böhning-Gaese, Katrin. 2002. Seed dispersal, breeding system, tree density and the spatial pattern of trees - a simulation approach. *Basic and Applied Ecology*, **3**(2), 115–123.

Bolker, Benjamin, & Pacala, Stephen W. 1997. Using Moment Equations to Understand Stochastically Driven Spatial Pattern Formation in Ecological Systems. *Theoretical Population Biology*, **52**(3), 179–197.

Canham, Charles D, LePage, Philip T, & Coates, K Dave. 2004. A neighborhood analysis of canopy tree competition: effects of shading versus crowding. *Canadian Journal of Forest Research*, **34**(4), 778–787.

Carlon, David B., & Olson, Richard Randolph. 1993. Larval dispersal distance as an explanation for adult spatial pattern in two Caribbean reef corals. *Journal of Experimental Marine Biology and Ecology*, **173**(2), 247–263.

Comita, Liza S., Muller-Landau, Helene C., Aguilar, Salomon, & Hubbell, Stephen P. 2010. Asymmetric Density Dependence Shapes Species Abundances in a Tropical Tree Community. *Science*, **329**(5989), 330–332.

Condit, Richard. 1998. *Tropical Forest Census Plots*. Berlin, Germany, and Georgetown, Texas: Springer-Verlag and R. G. Landes Company.

Condit, Richard, Sukumar, R, Hubbell, Stephen P, & Foster, Robin B. 1998. Predicting population trends from size distributions: A direct test in a tropical tree community. *The American Naturalist*, **152**(4).

Condit, Richard, Ashton, Peter S., Manokaran, N., LaFrankie, James V., Hubbell, Stephen P., & Foster, Robin B. 1999. Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50ha plots. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **354**(1391), 1739–1748.

Condit, Richard, Ashton, Peter S., Baker, Patrick, Bunyavejchewin, Sarayudh, Gunatilleke, Savithri, Gunatilleke, Nimal, Hubbell, Stephen P., Foster, Robin B., Itoh, Akira, LaFrankie, James V., Lee, Hua Seng, Losos, Elizabeth, Manokaran, N., Sukumar, R., & Yamakura, Takuo. 2000. Spatial Patterns in the Distribution of Tropical Tree Species. *Science*, **288**(5470), 1414–1418.

Dale, Mark R. T. 1999. *Spatial pattern analysis in plant ecology*. Cambridge: Cambridge University Press.

Feeley, Kenneth J., Davies, Stuart J., Noor, Md. Nur Supardi, Kassim, Abdul Rahman, & Tan, Sylvester. 2007. Do current stem size distributions predict future population changes? An empirical test of intraspecific patterns in tropical trees at two spatial scales. *Journal of Tropical Ecology*, **23**(02), 191–198.

Feeley, Kenneth J., Davies, Stuart J., Perez, Rolando, Hubbell, Stephen P., & Foster, Robin B. 2011. Directional changes in the species composition of a tropical forest. *Ecology*, **92**(4), 871–882.

Felsenstein, Joseph. 1975. A Pain in the Torus: Some Difficulties with Models of Isolation by Distance. *The American Naturalist*, **109**(967), 359–368.

Getzin, Stephan, Wiegand, Thorsten, Wiegand, Kerstin, & He, Fangliang. 2008. Heterogeneity influences spatial patterns and demographics in forest stands. *Journal of Ecology*, **96**, 807–820.

Harms, Kyle E., Condit, Richard, Hubbell, Stephen P., & Foster, Robin B. 2001. Habitat Associations of Trees and Shrubs in a 50-Ha Neotropical Forest Plot. *Journal of Ecology*, **89**(6), 947–959.

He, Fangliang, & Duncan, Richard P. 2000. Density-dependent effects on tree survival in an old-growth Douglas fir forest. *Journal of Ecology*, **88**, 676–688.

Hubbell, Stephen P. 1979. Tree Dispersion, Abundance, and Diversity in a Tropical Dry Forest. *Science*, **203**(4387), 1299–1309.

Hubbell, Stephen P., Foster, Robin B., O'Brien, S. T., Harms, K. E., Condit, Richard, Wechsler, B., Wright, S. J., & de Lao, S. Loo. 1999. Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science*, **283**(5401), 554–557.

Hubbell, Stephen P., Condit, Richard, & Foster, Robin B. 2005. *Barro Colorado Forest Census Plot Data*. [online] https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Illian, Janine, Penttinen, Antti, Stoyan, Helga, & Stoyan, Dietrich. 2008. *Statistical analysis and modelling of spatial point patterns*. West Sussex, England: John Wiley.

Janzen, Daniel H. 1970. Herbivores and the Number of Tree Species in Tropical Forests. *The American Naturalist*, **104**(940), 501–528.

Karlson, Ronald H., Cornell, Howard V., & Hughes, Terence P. 2007. Aggregation influences coral species richness at multiple spatial scales. *Ecology*, **88**(1), 170–177.

Kenkel, N. C. 1988. Pattern of Self-Thinning in Jack Pine: Testing the Random Mortality Hypothesis. *Ecology*, **69**(4), 1017–1024.

Kress, W. John, Erickson, David L., Jones, F. Andrew, Swenson, Nathan G., Perez, Rolando, Sanjur, Oris, & Bermingham, Eldredge. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*, **106**(44), 18621–18626.

Law, Richard, Murrell, David J., & Dieckmann, Ulf. 2003. Population growth in space and time: spatial logistic equations. *Ecology*, **84**(1), 252–262.

Levin, Simon A. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, **73**(6), 1943–1967.

Levine, Jonathan M., & Murrell, David J. 2003. The community-level consequences of seed dispersal patterns. *Annual review of ecology, evolution and systematics*, **34**, 549–574.

Marshall, Roger J. 1991. A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **154**(3), 421–441.

Muller-Landau, Helene C., & Hardesty, Britta D. 2005. Seed dispersal of woody plants in tropical forests: concepts, examples and future directions. *Pages 267–309 of:* Burslem, Pinard, & Hartley (eds), *Biotic interactions in the Tropics*. Cambridge: Cambridge University Press.

Murrell, David J. 2009. On the emergent spatial structure of size-structured populations: when does self-thinning lead to a reduction in clustering? *Journal of Ecology*, **97**, 256–266.

Murrell, David J., Purves, Drew W., & Law, Richard. 2001. Uniting pattern and process in plant ecology. *Trends in Ecology & Evolution*, **16**(10), 529–530.

Pagel, Mark. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, **26**(4), 331–348.

Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature*, **401**(6756), 877–884.

Pocock, Michael J. O., Hartley, Stephen, Telfer, Mark G., Preston, Christopher D., & Kunin, William E. 2006. Ecological correlates of range structure in rare and scarce British plants. *Journal of Ecology*, **94**, 581–596.

Seabloom, Eric W., Bjørnstad, Ottar N., Bolker, Benjamin M., & Reichman, O. J. 2005. Spatial signature of environmental heterogeneity, dispersal, and competition in successional grasslands. *Ecological Monographs*, **75**(2), 199–214.

Seidler, Tristram G., & Plotkin, Joshua B. 2006. Seed Dispersal and Spatial Pattern in Tropical Trees. *PLoS Biology*, **4**(11), e344.

Sheather, Simon J. 2008. *A Modern Approach to Regression with R*. Springer Texts in Statistics. Dordrecht: Springer.

Wang, Xugao, Ye, Ji, Li, Buhang, Zhang, Jian, Lin, Fei, & Hao, Zhanqing. 2010. Spatial distributions of species in an old-growth temperate forest, northeastern China. *Canadian Journal of Forest Research*, **40**(6), 1011–1019.

Wassermann, Larry. 2006. *All of nonparametric statistics*. New York: Springer.

Wiegand, T., Gunatilleke, S., & Gunatilleke, N. 2007. Species Associations in a Heterogeneous Sri Lankan Dipterocarp Forest. *American Naturalist*, **170**(4), E77–E95.

Wilson, Robert J., Thomas, Chris D., Fox, Richard, Roy, David B., & Kunin, William E. 2004. Spatial patterns in species distributions reveal biodiversity change. *Nature*, **432**, 393–396.

Wright, S. Joseph, Hernandz, Andrs, & Condit, Richard. 2007. The Bushmeat Harvest Alters Seedling Banks by Favoring Lianas, Large Seeds, and Seeds Dispersed by Bats, Birds, and Wind. *Biotropica*, **39**.

Chapter 4

Detecting sub-communities in ecosystems from multivariate spatial associations

Summary

Species are seldom distributed across an ecosystem at random, but instead show spatial structure that is determined by environmental gradients and/or biotic interactions. Previous studies have focused on univariate distributions of species or pairwise associations between two species to investigate the effect of environmental factors, biotic interactions, or species traits on the spatial arrangement of individuals. We propose a multivariate method which uses the spatial co-associations between all pairs of species to find sub-communities of species whose distributions in the study area are positively correlated. We use the sub-communities to construct a map of the spatial structure of the ecosystem, which can then be analysed to explore the effect of ecological processes that may have caused the spatial structuring. Our method is particularly well-suited for ecosystems with large numbers of species and gives rare species a strong weight.

Using data on the distribution of tree and shrub species from a 50 ha forest plot on Barro Colorado Island (BCI), Panama, we show that our method can be used to construct biologically meaningful maps of the spatial structure of the ecosystem. In particular, we detect habitats based on environmental gradients (such as slope) as well as different biotic conditions (such as canopy gaps) and species groups with similar biological traits (shade tolerance). We discuss extensions and adaptations to our method that might be appropriate for other types of spatially referenced data and for other ecological communities. We make suggestions for other ways to interpret the sub-communities using phylogenetic relationships, biological traits, and environmental variables as covariates. We also note that sub-communities that are hard to interpret may suggest groups of species and/or regions of the landscape that warrant further attention.

Impact of this work

With large data-sets of species rich communities becoming increasingly available, better methods to explore multivariate structure are increasingly important. Abundance and within species spatial patterns have a strong impact on the expected between species spatial patterns of pairs of species, which explains why spatial co-associations between different pairs of species are difficult to compare. In this chapter we introduce a new method to normalise pair-wise co-association values to make them comparable. We then show how the spatial structure of a multi-species community can be analysed based on that normalised co-association matrix of between species spatial co-associations.

Declaration on the contributions to the work presented in this chapter:

This chapter is submitted as a manuscript for publication and currently under review:

Detecting sub-communities in ecosystems from multivariate spatial associations, Anton J. Flügge, Sofia C. Olhede, and David J. Murrell, under review, 2013.

The normalisation method was developed by me. All analyses were conducted by me. The first draft was written by me. Sofia Olhede and David Murrell contributed by supervising and guiding my work and by revising the manuscript before submission for publication. The data used was provided by the Centre for Tropical Forest Science.

4.1 Introduction

Understanding the processes that underpin observed patterns of biodiversity and how functionally similar species co-exist in close spatial proximity are among the primary challenges in ecology (Hardin, 1960; Wright, 2002). Biodiversity may bolster ecosystem stability and productivity (Isbell *et al.*, 2009; Cardinale *et al.*, 2012), and is therefore an important aspect in the environmental services that ecosystems provide to society. However, land-use changes (Brooks *et al.*, 2002), growing populations (Williams, 2012), and climate change (Bellard *et al.*, 2012) may all threaten biodiversity and ecosystems in general. Understanding the ecological processes that both create and maintain high biodiversity is important for protecting diverse ecosystems.

Through organisations like the Center for Tropical Forest Science (CTFS, 2013) there is now data available of multiple large-scale forest plots for which all trees and shrubs are individually mapped and identified to species level. Most spatial analyses of these spatially referenced individual-based tree datasets have been univariate, investigating the spatial distributions of different species and looking for links between the within species spatial pattern to other processes/factors such as abundance (Condit *et al.*, 2000); recent changes in local abundance (Flügge *et al.*, 2012); dispersal mechanism (Muller-Landau & Hardesty, 2005);

conspecific density dependence (Bagchi *et al.*, 2011); and habitat association (Harms *et al.*, 2001; Ledo *et al.*, 2013; Itoh *et al.*, 2010). Some studies have begun to consider pairs of species to investigate the effect of species interactions (e.g. Wiegand *et al.*, 2012), but few studies have considered a multivariate approach where the spatial co-associations of all pairs of species are jointly considered.

A multivariate approach is useful because it will highlight any groups of species that are found together more often than expected by chance, and once these groups of species have been identified it is possible to investigate the processes that are driving their spatial association. Theory has shown that strong interspecific competition should lead to negative spatial associations as heterospecific individuals are removed from neighbourhoods (Murrell *et al.*, 2001). On the other hand, positive spatial associations can occur if species interactions are positive (Callaway, 1995) or if species have shared preferences in habitat. It may even be possible that interspecific clustering occurs if two weakly competing species share natural enemies, although we note there is less theory on this subject. Dispersal limitation, a purely stochastic process, may also lead to some strong positive or negative associations, but overall one would expect it to create spatial independence between species, and this represents the null model.

In what follows we outline a method for grouping together species according to their interspecific spatial associations, and we highlight the potential of this approach with an example where the interpretation of the groups of species is based upon environmental niches. However, the reader should note that other datasets may require different interpretations (perhaps based upon traits or competition) and our main focus here is on presenting a method of grouping species together in a meaningful way. The method has three steps. First the interspecific associations need to be quantified, taking into account differences in abundance and within species spatial distribution. The second step involves using an algorithm to group species together that have similar spatial co-associations. The final step is to then create a map denoting locations in the landscape where each sub-community dominates. To illustrate our approach we use the Barro Colorado Island (BCI) forest dynamics plot (Hubbell *et al.*, 2005), which allows us to contrast our results with those of previous studies. In particular, we compare our results to previous work on habitats at BCI (Harms *et al.*, 2001; Kanagaraj *et al.*, 2011).

The role of environmental niches in species co-existence is a particularly contentious aspect. While niche theory (Tilman, 2004) proposes that the co-existence of multiple species is supported by differences in their tolerance to, and preference for, specific environmental and biotic conditions, neutral theory (Hubbell, 2001) highlights random speciation, survival and dispersal events as the basis for many of the patterns of biodiversity observed by ecologists. Though both niche and neutral processes are likely to contribute to species richness, the relative importance of each is difficult to determine (Adler *et al.*, 2006). This is due to at

least two factors, each of which, we aim to address in this study. First, past attempts to look for effects of the environment on the distributions and diversity of species, have relied on human experts. They were based on expert knowledge regarding which topographically defined regions (Harms *et al.*, 2001) or environmental variables (Itoh *et al.*, 2010; Kanagaraj *et al.*, 2011) describe meaningful habitats potentially structuring the niche space for species. This approach has the danger to miss niche-effects that are due to environmental inhomogeneities not obvious to human observers (Ledo *et al.*, 2013). In addition the choice of environmental variables or spatial region might implicitly already be influenced by the observed spatial patterns of species which invalidates statistical tests for spatial correlations. Second, in understanding patterns of biodiversity, rare species are of great interest. Most communities exhibit a skewed distribution of population abundances/densities such that the majority of species are represented by relatively few individuals, and this effect is often most pronounced in species-rich communities such as rain forests (Hubbell, 2001). Therefore, ignoring the rarer species risks throwing away valuable information. Moreover, common species may be the most tolerant with respect to the range of environmental conditions (i.e., habitat generalists), while rare species are likely to be most restricted by their niche breadth (Kanagaraj *et al.*, 2011). However, studying rare species is plagued with difficulties. When studying individual species, statistical power is normally lowered as the species become rarer, and abundance thresholds are imposed to retain only those species where the signal to noise ratio is likely to be high (e.g. Harms *et al.*, 2001). On the other hand when species are pooled together the impact of individual species on the overall result can be dependent on their abundance (e.g. Kanagaraj *et al.*, 2011), which is giving common species much more weight than rare species.

Harms *et al.* (2001) conducted a study on the effects of habitat on the spatial patterns of woody plants at Barro Colorado Island (BCI) in Panama. The authors manually chose threshold values for environmental variables, such as slope, altitude, and water availability, to delineate five different habitat types (excluding a "young forest" habitat based on known forest history, i.e., where the forest had been cleared in the past). They then investigated whether each species of tree or shrub could be positively or negatively correlated with each of the five pre-defined habitats. The results showed that only 171 out of 855 possible interactions between species and habitat could be reliably classified as non-random. In a separate study of the BCI plot, Kanagaraj *et al.* (2011) used multivariate regression trees to automatically find environmental variables and threshold values that partition the landscape into different habitats based on the similarity of the species composition found in 20x20 m patches. For the population of juveniles, this method largely provided support for the habitats defined by Harms *et al.* (2001). However, in reproductive adults, the effect of habitat disappeared, which the authors interpreted as support for the neutral theory. Similar studies testing for habitat associations of species with pre-defined habitats have been conducted at other study

sites (e.g. Valencia *et al.*, 2004; Ledo *et al.*, 2013). It is not clear that the pre-defined habitats or the environmental variables chosen for partitioning the space are the most relevant for most species; naturally, these methods can only work according to the assumptions that are fed into the analyses. In contrast to those previous studies, using our method, we refrain from using any a priori knowledge on the environment or any assumptions on which spatial region or environmental condition constitute a habitat. Instead we start from the spatial distribution patterns of the species in the community, and look for structure in their spatial correlations.

4.2 Materials and methods

4.2.1 Data used from Barro Colorado Island (BCI)

We use the data from the Barro Colorado Island (BCI) 50 ha long-term forest dynamics plot in Panama (see Condit, 1998; Hubbell *et al.*, 1999, 2005). The forest plot at BCI was established in 1980 and from 1985 onward, complete censuses of all trees and shrubs above 1 cm diameter at breast height (DBH) were repeated every five years. All individuals are identified to species level and their position and size is recorded in every census. Each individual is classified as adult or juvenile by comparing it to a species-specific DBH criteria based on estimates by Robin Foster on the typical sizes when species become reproductive (R. Foster, unpublished data). Our analysis includes 141 shrub and tree species (out of 301 species), namely those with at least 10 adults and 10 juveniles in the most recent 2010 census. Species with fewer individuals are excluded because in this instance it is not possible to estimate reliable co-association values for both the juvenile and the adult populations. This criterion excludes very rare species, species that do not reproduce in the plot itself, and small shrub species for which all individuals included in the census are classified as adults. Considering only adult niche requirements would lead to the inclusion of another 34 species, 26 of which are shrub species. In our interpretation of the results we also use the shade-tolerance indices (available for 124 of the 141 species) from Comita *et al.* (2010). In total the analyses that follow include 153,634 trees and shrubs (35,156 adults and 118,478 juveniles) out of 207,259 individuals above 1 cm DBH in the 2010 census (see Appendix 7.2 for list of species, abundances, shade-tolerance indices and Robin Foster estimates).

4.2.2 The normalised co-association matrix and the sub-community maps

4.2.2.1 Bivariate co-association measure

As a measure of the spatial co-association of two species, we use the bivariate version of the $\Omega_{0,10}$ relative neighbourhood density as defined in Condit *et al.* (2000). The relative neighbourhood density $\Omega_{0,10}$ is similar to the cross-pair correlation function (Law *et al.*, 2009), and is proportional to the widely-used Ripley's K (Wiegand & Moloney, 2004), but is standardised for the area of the neighbourhood that is analysed. This means the expected

value of the bivariate $\Omega^{(a,b)}$ for a random superposition is equal to one, independent of scale. The bivariate $\Omega_{0,10}^{(a,b)}$ relative neighbourhood density counts how many individuals of species b can be found in an average neighbourhood of 10 m around an individual of species a and is defined as:

$$\Omega_{0,10}^{(a,b)} \equiv \frac{A \sum_{i=1}^{N^a} N_{0,10}^{(a_i,b)}}{N^b N^a A_{0,10}}, \tag{4.1}$$

where $N_{0,10}^{(a_i,b)}$ is the number of neighbours of species b within the interval 0 to 10 m from a focal individual i of species a; N^a and N^b are the total number of individuals of the respective species in the sample; and $A_{0,10}$ is the size of the area of a 10 m circle around the focal trees. Edge effects are corrected for by using a buffer zone which was created by excluding individuals of species a from the sample that were closer than the neighbourhood radius to the edge of the study area (Haase, 1995).

4.2.2.2 Co-association matrix and normalisation

It is necessary to normalise $\Omega_{0,10}^{(a,b)}$ because the tree and shrub species vary both in their abundance and in their within species spatial association. Consequently, it is difficult to compare the co-association measures defined above for different pairs of species in a meaningful way. We therefore normalise the co-association values $\bar{\Omega}_{0,10}^{(a,b)}$, accounting for marginal within species aggregation, where we define:

$$\bar{\Omega}_{0,10}^{(a,b)} \equiv \begin{cases} 0 & \text{for } a = b \\ \frac{\Omega_{0,10}^{(a,b)} - 1}{\text{std}(\mathbf{\Omega}_{\mathbf{0,10}}^{\mathbf{R(a,b)}})} & \text{for } a \neq b \end{cases}. \tag{4.2}$$

The quantity $\mathbf{\Omega}_{\mathbf{0,10}}^{\mathbf{R(a,b)}}$ is the vector of 1000 $\Omega_{0,10}^{(a,b^*)}$ values which are computed by random torus translations (Harms *et al.*, 2001) of the spatial locations of species b in relation to the spatial locations of species a. We remove unity from Equation 4.2 because under the assumption of random superposition (spatial independence) of two species the expectation of $\Omega_{0,10}^{(a,b)}$ is unity. By doing so, we shift the co-association values such that, compared to a null model of random superposition, negative values indicate co-segregation and positive values indicate co-aggregation. Our normalisation procedure uses resampling methods, in order to keep the marginal distribution of species constant and to avoid confounding the effects of abundance and within species aggregation with the bivariate $\Omega_{0,10}^{(a,b)}$. As $\Omega_{0,10}^{(a,b)}$ is identical to $\Omega_{0,10}^{(b,a)}$, except for an asymmetry in the estimation introduced by the edge correction which does not affect the expected value, it is sufficient to compute the upper or lower triangle of the matrix to obtain the symmetric matrix of all pairwise co-association values. The diagonal entries of the matrix are set to zero, because we are not interested in the within species spatial associations.

4.2.2.3 Clustering of species into sub-communities

We use the popular non-hierarchical k-means clustering algorithm (Gan *et al.*, 2007) to group the species into k disjunct sets of species with the most similar co-association values. Species are represented by the rows of the normalised co-association matrix. We use 100 replications of the k-means algorithm with random initialisation to find the clustering that minimise the sum of the difference $D(k)$ between the vectors of co-association values of species and the centroid of their cluster. We define $\bar{\mathbf{\Omega}}_{0,10}^{(x,\cdot)}$ as the vector of normalised co-association values between all species and species x, and c_x as the cluster number of the cluster to which species x is assigned. The sum of the difference $D(k)$ is then computed as the sum of the differences of each species co-association vector $\bar{\mathbf{\Omega}}_{0,10}^{(x,\cdot)}$ to the cluster center of its cluster c_x:

$$D(k) = \sum_{i \in \text{Species}} \left\| \bar{\mathbf{\Omega}}_{0,10}^{(i,\cdot)} - \frac{\sum_{j \in \text{Species}} \delta_{c_i,c_j} \bar{\mathbf{\Omega}}_{0,10}^{(j,\cdot)}}{\sum_{j \in \text{Species}} \delta_{c_i,c_j}} \right\|_2 \tag{4.3}$$

with the Kronecker delta, δ_{c_i,c_j}, defined to be zero if $c_i \neq c_j$ and one if $c_i = c_j$.

The result is that each group is a collection of species that are most aggregated in the same areas of the 50 ha plot and are most similarly segregated from the other species. To determine the upper limit of k for which the individual clusters contain meaningful information on the spatial patterns of the species, we use the normalised co-association matrix for 1000 random forests in which the within species pattern is held constant, but where all species are shifted relative to each other via random torus translations. For both the BCI data and each random forest, we then compute the sum of the within-cluster species to centroid distances $D(k)$ for all k between 1 and the number of species. The sum of the difference $D(k)$ is a measure of how well the clustering fits the data (i.e., of how homogeneous the species are within a cluster). With increasing k, $D(k)$ trivially gets smaller, because more clusters can always partition a set such that the sum of the within cluster distances is smaller than with fewer clusters. However, the amount by which $D(k)$ decreases from $D(k_1)$ to $D(k_2)$ (with $k_2 > k_1$) holds information on the inherent number of clusters in the data. We therefore compare $D(k) - D(k+1)$ for all k between 1 and 140 (number of species minus 1) for the BCI data with the 1000 random forests. If the species at BCI are more likely to be found in the same or different spatial regions, we would expect $D(k) - D(k+1)$ to be larger than in a random forest for at least the first few clusters k. This would show that the structure of the forest is not random, but that there are indeed sub-communities that reduce the sum of within cluster distances more than what would be expected in a random null-model. To avoid interpreting potentially spurious effects, at most those number of clusters k that exhibit statistical significance are investigated. In the analyses below we use a 1% significance level, i.e., k is significant if $D(k) - D(k+1)$ for the species at BCI is larger than for 99% of the random forests.

4.2.2.4 Density maps

Once all species are grouped, the next step is to explore the spatial distribution of the sub-communities in the landscape. For that purpose we first use the kernel density estimator by Botev *et al.* (2010) to estimate the relative density of each species across the 50 ha plot. We then compute the mean relative density across the 50 ha for each sub-community. In line with our clustering method, this method of computing the relative density for sub-communities weighs each species identically, independent of its abundance. As we have argued in the introduction, rare species might show stronger habitat preferences, while more abundant species might be more likely to be generalists; hence, if we are interested in habitats we should not overlook rare species (Kanagaraj *et al.*, 2011). Also we are less concerned with the absolute density of individuals in a certain region (in which case we should weigh the species density maps by abundance or basal area), but instead we want to find regions at which most species in the sub-community co-occur. Estimates of species density are less reliable for rare species which could be another reason to weigh sub-community density maps by abundance, but Botev *et al.*'s (2010) kernel density estimator uses a variable bandwidth that adapts to the level of detail available in the data, and smooths the species density maps over larger scales for rare species, taking into account the lower reliability of the data.

4.2.2.5 Sub-community maps

The information in the sub-community density plots can be condensed into a single panel showing the dominant cluster, i.e. the sub-community with the highest mean relative density, for each 20-by-20 m quadrant in the forest plot. Below, we draw such a figure by representing each sub-community with a different colour, and drawing a map of the 50 ha forest plot where each 20-by-20 m quadrant is coloured according to the sub-community that has the highest mean relative density.

4.3 Results

4.3.1 Co-association matrix and normalisation

Figure 4.1 shows the normalised co-association matrix of the adult population in the 2010 census. Each row (and column) represents the co-association values of one species with all others. The colour coded bars along the side of the matrix show the clustering of the species for $k = 5$ clusters (the colors are the same as those in Figure 4.5d).

4.3.2 Clustering of species into sub-communities

Comparing $D(k) - D(k+1)$ between the data of the adult individuals of the 141 study species at the 2010 census, and 1000 random forests based on the same individuals, shows that $D(k) - D(k+1)$ is larger for the true data than for 99% of the random forests up to values of $k = 10$ (see Figure 4.2 and Figure 4.3). This indicates that at least ten disjoint sets of species can be defined on the basis of their spatial distribution within the BCI forest plot,

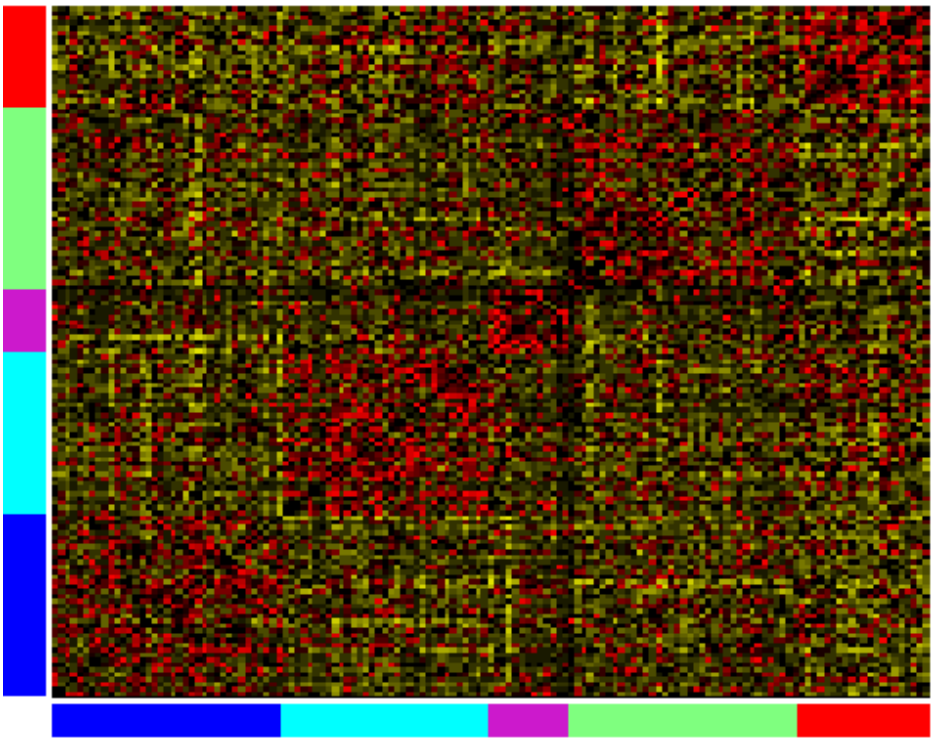


Figure 4.1: The normalised co-association matrix between the adults of the 141 tree and shrub species from the BCI plot investigated here. Each row and column represents the co-associations of one species with all the others. Red indicates that two species are aggregated, and yellow that they are segregated, in comparison to a random null-model. The matrix is symmetric about the diagonal, and the colors on the side show which species are grouped together in one sub-community by the k-means algorithm (with $k = 5$) using the same color-coding as in Figure 4.5d.

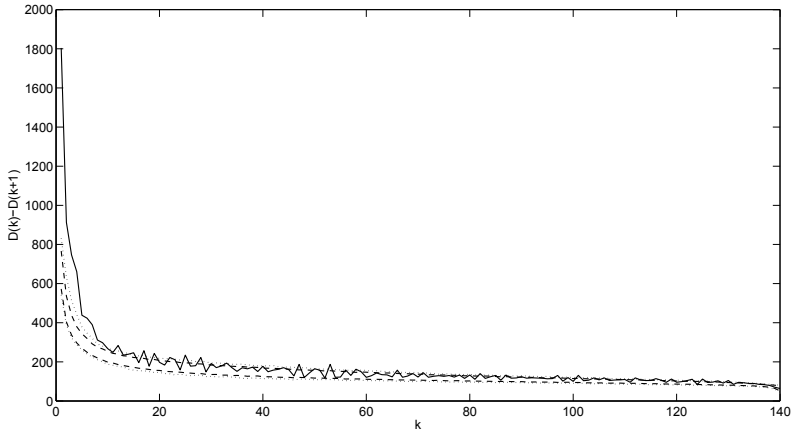


Figure 4.2: Lines show the values of D(k)-D(k+1) (as defined in Section 4.2.2.3). The solid line shows the values for the true data for adults of the 141 species in the analysis, while the dotted lines show the 1st and 99th percentile and the dashed line the 5th and 95th percentile of 1000 randomised forest structures.

Sub-community name	Number of species	Number of adults	mean shade-tolerance (±std)
South-eastern low plateau	37	7821	0.55 (±0.93)
North-western low plateau	33	2488	-0.63 (±1.48)
Swamp/Shade-intolerant pioneers	13	1997	-1.57 (±1.96)
High plateau/Young forest	37	17164	0.17 (±0.95)
Slope	21	5686	0.45 (±0.70)

Table 4.1: Summary information on the number of species, number of adult individuals, and mean shade-tolerance index for the clustering with $k = 5$ sub-communities on the basis of the adult individuals in the 2010 census at Barro Colorado Island (BCI).

and within these sets species are more correlated than expected by chance, i.e., assuming spatial independence between pairs of species. For $k > 10$, however, the additional fine scale structure in the data can no longer be distinguished from random effects. In order to enable better comparison of our results to those obtained by Harms *et al.* (2001), we choose to use $k = 5$ for most of our analyses (see Table 4.1 and 4.2 for summary information on the clustering with $k = 5$). The comparison with the random forests shows with very high confidence that for $k = 5$, the structure we find is the product of non-random spatial processes.

Sub-community name	Number of species	Number of juveniles	mean shade-tolerance (±std)
Blue/low plateau	32	41573	0.95 (±0.51)
Light blue/Mixed	37	10498	-0.01 (±0.98)
Purple/Shade-intolerant pioneers	13	3720	-2.51 (±1.34)
Green/Mixed-swamp	26	17448	-0.56 (±1.33)
Red/Slope	33	41573	0.49 (±0.50)

Table 4.2: Summary information on the number of species, number of juvenile individuals, and mean shade-tolerance index for the clustering with $k = 5$ sub-communities on the basis of the juvenile individuals in the 2010 census at Barro Colorado Island (BCI).

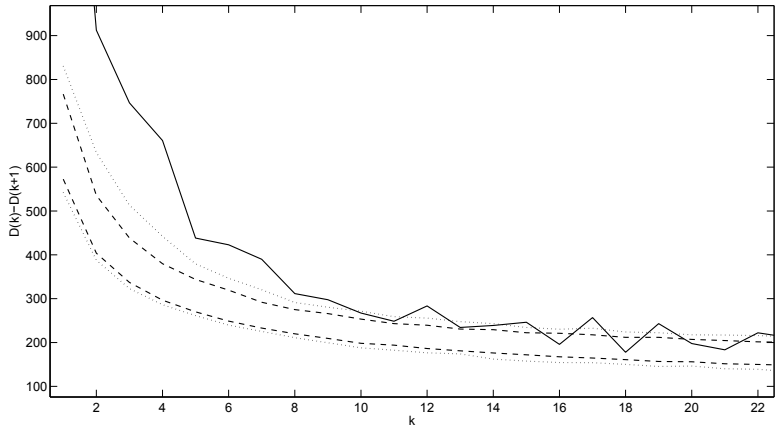


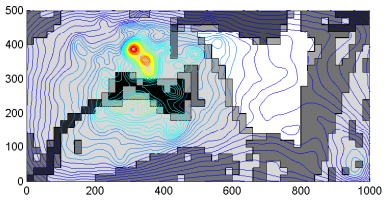
Figure 4.3: Close-up of Figure 4.2. Lines show the values of D(k)-D(k+1) (as defined in Section 4.2.2.3). The solid line shows the values for the true data for adults of the 141 species in the analysis, while the dotted lines show the 1st and 99th percentile and the dashed line the 5th and 95th percentile of 1000 randomised forest structures..

4.3.3 Density maps

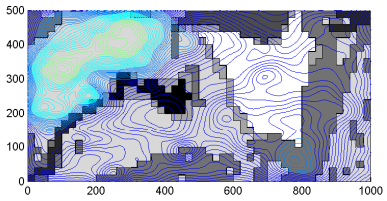
The panels of Figure 4.4 show the mean density of the adult individuals of each of the five sub-communities as estimated by Botev's kernel density estimator (Botev *et al.*, 2010).

4.3.4 Sub-community maps

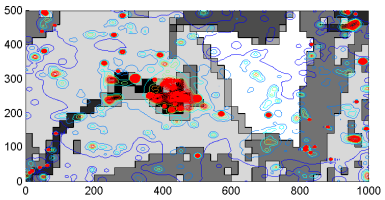
Figure 4.5 shows the dominant sub-communities in the 50 ha forest plot for k between 2 and 5 clusters for the adult population of the 2010 census (see Appendix 7.3 for sub-community maps for number of clusters up to $k = 10$). Our results largely concur with those of Harms *et al.* (2001). The first partitioning at $k = 2$ (Figure 4.5a) seems to distinguish between the more wet habitat at the slopes and the drier plateau habitats. However, at $k = 2$, part of the north-western low plateau is grouped together with the slope (colored red), rather than with the remainder of the plateau habitat (colored green). For $k = 3$ (Figure 4.5b) we do not find a distinction between the high plateau and the low plateau. Instead, the partitioning follows similar borders as the first, except that the north-western low plateau (cyan) stands out as a separate sub-community. Thus, the remaining parts of the low plateau are still grouped together with the high plateau and the young forest. Only when increasing the number of clusters to $k = 4$ (Figure 4.5c), do we find a separate high plateau/young forest sub-community while still finding the split between the north-western and the south-eastern part of the low plateau. For $k = 5$ clusters (Figure 4.5d, based on the sub-community densities shown in Figure 4.4), we find a sub-community dominated by swamp species, together with some more widely spread shade-intolerant pioneer species. The divide between the north-western low plateau and the south-eastern low plateau seems to be mainly driven by life history strategy, since the species of the south-eastern sub-community have the highest mean shade-tolerance index of all clusters (Table 4.1), and the species of the north-western low plateau



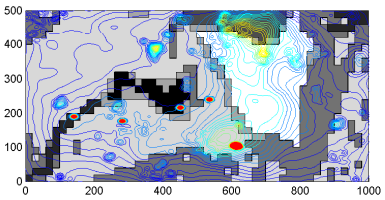
(a) South-eastern low plateau



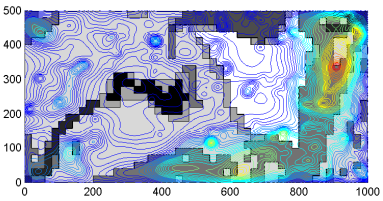
(b) North-western low plateau



(c) Swamp/Shade-intolerant pioneers

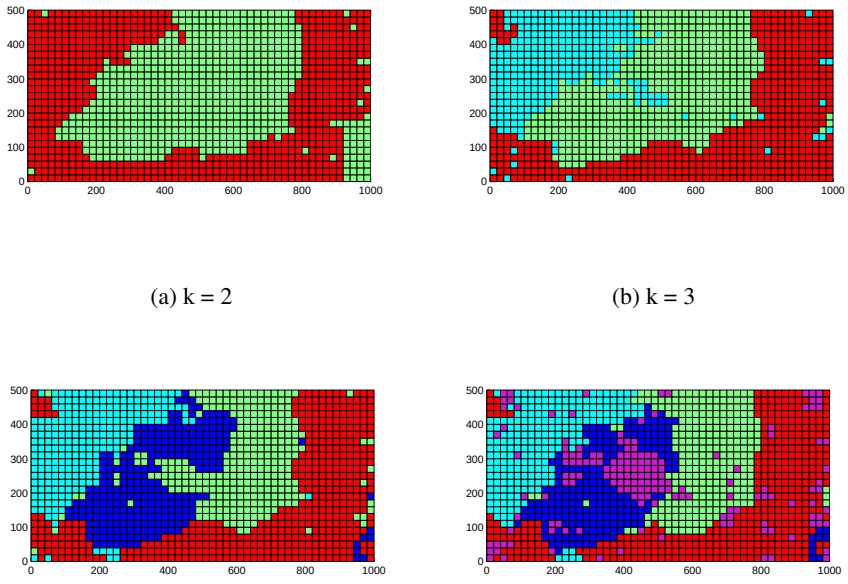


(d) High plateau/Young forest



(e) Slope

Figure 4.4: Each panel shows the mean relative density of adults in the 2010 census obtained for one of five sub-communities. Red indicates that there are many individuals from that set of species while blue indicates lower densities. Densities were computed using Botev *et al.*'s (2010) kernel density estimator for each individual species and then averaged over all species in each sub-community. The gray-scale map in the background shows the different habitats at Barro Colorado Island (BCI) as defined by Harms *et al.* (2001).



(a) k = 2 (b) k = 3

(c) k = 4 (d) k = 5

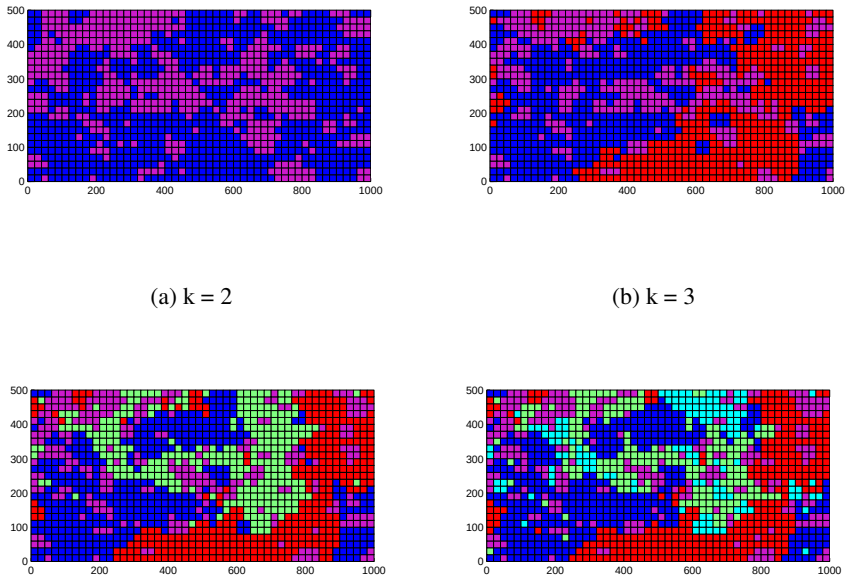
Figure 4.5: Depicted are the sub-communities with the highest mean relative density for each 20-by-20 m quadrant for the number of cluster k between $k = 2$ and $k = 5$ (top-left to bottom-right) for the adult plants in the 2010 census.

sub-community are the second most shade-intolerant on average (only the swamp/pioneer sub-community has a lower mean shade tolerance index).

The results for the juveniles (Figure 4.6) are slightly less clear, although life-history strategy seems to be an important factor differentiating the sub-communities, suggesting light gaps drive some of the spatial structure evident in the plot. Most notably, the first grouping for $k = 2$ (Figure 4.6a) seems to be made along the line of shade tolerance (mean shade-tolerance index for the "purple" sub-community is -1.01 ± 1.55; for the "blue" sub-community, it is 0.50 ± 0.79). For $k > 2$ (Figure 4.6b – d), there always seems to be a sub-community of highly shade intolerant species beside those sub-communities that are more influenced by habitat and more similar to the sub-communities found for the adults (see Table 4.2 for summary information on the juveniles with $k = 5$). The results from the juveniles support the result from the adults that slope is the most important environmental variable to distinguish habitats with different species compositions at BCI.

4.4 Discussion

There are an increasing number of data sets available that provide rich spatial data of many species (CTFS, 2013). Most analyses have so far focused on the spatial distributions of individual species, only aggregating the results to summarize the number of species that show certain spatial associations (e.g. Harms *et al.*, 2001) or reporting a median value (e.g. Condit



(a) k = 2 (b) k = 3

(c) k = 4 (d) k = 5

Figure 4.6: Depicted are the sub-communities with the highest mean relative density for each 20-by-20 m quadrant for the number of clusters k between $k = 2$ and $k = 5$ (top-left to bottom-right) for the juvenile plants in the 2010 census.

et al., 2000). There have been fewer attempts to draw information from the joint spatial pattern of all species and yet this may yield additional insight into the processes that dominate the communities. The method we introduce makes use of spatial co-association measures to group species together based upon their co-occurrence in the landscape, and we believe the interpretations of the groups of species should be used in addition to the more common univariate approach of considering within species spatial structure. In what follows we will first discuss extensions and adaptations of our method that may be required for different datasets before then discussing the results of our example.

The first step is to calculate a normalised co-association matrix. The matrix of co-associations is itself an interesting object that could be used for other analyses such as the comparison of the degree of segregation and aggregation between different groups of species or ecosystems (see also Chapter 5). The normalisation procedure we outline is necessary to make the individual entries of the matrix comparable, but the precise measure that is used to compute co-associations will differ depending on the scale of the processes of interest in the particular ecosystem and the available data. We use circles of 10 m diameter to compute co-association values for our analyses, as this is a scale at which many important ecological processes are happening in our study system (Uriarte *et al.*, 2004). This scale also provides a good balance between covering a wide enough area to achieve stable numerical results (i.e., for most species pairs, we find individuals of the other species in at least some of the neigh-

bourhoods around the focus species) while still capturing fine-scaled differences in species distribution (Flügge *et al.*, 2012). Different species may show spatial correlations at different spatial scales, but because we were looking at the spatial structure of ecosystem with all species, we used the same scale for all species pairs to keep the results comparable. The results are relatively stable to changes of the diameter of neighbourhoods (see Appendix 7.4 for results on the 5 m and 20 m scale) as the neighbourhood densities at different scales are highly correlated (Condit *et al.*, 2000), and species that are assigned to a different cluster at a different scale, are likely to be the least typical species for that sub-community anyway. Although individuals in the BCI plot have their precise x- and y-coordinates recorded, the method could easily be adapted if the data is for presence/absence in a grid since the key ingredient required to group the species is a matrix of spatial co-associations.

The second step defines sub-communities from the co-association matrix and explores the number of statistically significant clusters in the data. Giving an upper limit for the number of sub-communities that can be distinguished from random is an important aspect of our method. However, the clustering method could be adapted. For example, if the expectation is that clusters break down into sub-clusters (e.g., groups relating to slope habitat break down into upper and lower slope groups), then hierarchical clustering methods (Gan *et al.*, 2007) could be used. Our results suggest there is a statistically significant structure in the co-association matrix for up to ten sub-communities for the adults at BCI (see Section 4.3.2), but we argue that below this cut-off, there is no a priori correct number of clusters. By considering different numbers of sub-communities we can explore which spatial structures show up first and are therefore the strongest. In our main analyses, we concentrated on $k = 5$ clusters because this allowed easy comparison with previous analyses using different methods. However, further in-depth analysis of the characteristics of the species in the $k = 10$ clusters could lead to new insights into ecologically important factors structuring the ecosystem at BCI, although more information about the species and/or the abiotic environment is likely to be required to explain larger numbers of clusters. A possible extension of our study could be to use the data on soil chemicals available for the BCI plot (Dalling *et al.*, 2009; Condit *et al.*, 2013). Baldeck *et al.* (2013) indicate that soil properties can explain a significant part of the spatial distribution of species at BCI, and we suspect this might explain the more fine-grained structure when k equals 6 to 10. We also note that instead of looking at density maps and sub-community maps one could also stop at the level of the clusterings and analyse the attributes of species in the various sub-communities to explore what they have in common and what distinguishes them (see Chapter 5). In this case the focus would be on species traits such as wood density, seed size, maximum adult size etc. (Wright *et al.*, 2010) or investigating the within and between sub-communities pattern of phylogenetic relatedness.

Although the interpretation of our species clusters focuses mainly on the role of environmental niches, other biotic and abiotic processes may be influential and the sub-communities

represent realised rather than fundamental niches. In our example, it is clear that sub-communities are influenced by both environmental gradients such as slope and elevation, but also biotic conditions such as canopy gaps caused by tree fall. The biotic factors may include both positive as well as negative forces acting on species spatial pattern since species are clustered according to similar positive and negative interspecific associations. A canopy gap for example provides particularly advantageous conditions for shade-intolerant species (Wright *et al.*, 2003). On the other hand shared pathogens or superior competitors could conceivably restrict the range of some species to those parts of the forest where the pathogen or competitor is not present. Spatial clustering of groups of species could also arise from positive interactions between the species. While positive interactions between the species in the BCI plot are thought to be rare (Volkov *et al.*, 2009; Peters, 2003; Wiegand *et al.*, 2012), they may be more important in other communities (e.g. Callaway *et al.*, 2002).

The third step takes the species clusters and computes relative density maps; a number of adaptations could be required depending on the data used. Firstly, we use a density estimation kernel that smooths the individual stem map to produce a continuous density distribution over the whole area. If the data was based on presence/absence in a grid then a different kernel density estimator would be appropriate. Secondly, because we were particularly interested in rare species and the co-associations between species we weight each species equally in the calculation of the mean sub-community density, but other methods of weighting the contribution of species depending on abundance, biomass or other measure of relevance or reliability of the data are possible. Weighting by stem abundance might bias the results towards species that produce lots of juveniles, whereas weighting by biomass would bias towards species that produce large individuals and both might lead to different, but biologically informative interpretations.

In the final step of our method we summarise the information from the density maps of the sub-communities into a single sub-communities map. This reduces the information from multiple individual maps into one for the human observer potentially easier to interpret summary map. We could, however, also have stopped at the density maps and for example analysed the correlation between the density of the various sub-communities to the value of other continuous variables, such as soil nutrients.

To highlight our multivariate method, we consider a community where we are able to include 141 species. For systems with fewer species the information gain compared to methods that focus on the spatial pattern of individual species is expected to be lower. However, there is no minimum number of species for our method to be applicable. If we have too little or noisy data the estimation of the number of clusters that are statistically different from random will indicate the limits of the analyses that is suitable.

As discussed above, our method has numerous possibilities and the example we give shows the potential of the method to describe habitats from the spatial associations of species.

We believe this is possible in the BCI dataset because previous research has suggested between species interactions are relatively weak and therefore less important in determining interspecific spatial associations (Comita *et al.*, 2010; Wiegand *et al.*, 2012). The role of environmental heterogeneity for the co-existence of high numbers of species has been a core interest of multiple recent studies (Harms *et al.*, 2001; Valencia *et al.*, 2004; Kanagaraj *et al.*, 2011; Ledo *et al.*, 2013; Punchi-Manage *et al.*, 2013). The approach commonly taken in these studies is to use the available information to first define habitats, or select relevant environmental variables based on human expert knowledge, and then use the spatial distribution of each species to look for correlations with those pre-defined habitats. The disadvantage of this approach is that the results depend heavily on the chosen habitats and the data available on that specific environment. Our method makes fewer assumptions concerning the spatial regions or environmental gradients along which the ecosystem is structured. The results show that our method is able to detect the major habitat types at BCI as defined by Harms *et al.* (2001) without using any prior information on the environment. In particular, we provide support for the distinct nature of the slope habitat in both adults and juveniles, but also, in the case of the adults for the swamp and the low and high plateau habitat (Figure 4.5). The low plateau, however, seems to be not entirely homogeneous. Our results suggest that the north-western part is distinct from the south-eastern part, something that could be caused by differing disturbance histories which may have lead to higher numbers of shade-intolerant species in the north-west. This could also be the reason why, despite the similar environmental conditions throughout the low plateau, Harms *et al.* (2001) found comparatively few species that were significantly correlated to that habitat (9 positively and 19 negatively associated out of a total of 171 species). We do not find a separate young forest sub-community for the region along the northern border of the plot that is known to have been cleared in the past (Condit, 1998), but most parts of the young forest are grouped together with the high plateau. This may indicate that the succession process has progressed sufficiently for this area to no longer differ from other parts of the high plateau. However, there are indications that some species are still found in much larger numbers in the young forest region (anonymous reviewer, personal communication). It is possible that habitats that are dominated or defined by only very few species are less likely to show up with k-means clustering because separating off these species into their group may not greatly decrease the within cluster distances, and this highlights the need to also consider individual species analyses.

One drawback of starting with pre-defined habitats and then using individual tree locations to determine niches is that the statistical power to find negative or positive associations of a species with a habitat depends on the size of the habitat and the abundance of the species. Smaller habitats have a smaller expected number of individuals and therefore a lower signal-to-noise ratio. Also, for large habitats, like the low plateau at the BCI plot (Figure 4.4 a, b), it can be difficult to find positive associations and easier to find negative associations. Con-

sider the extreme case where the landscape is almost entirely one habitat except for a very small region in one corner; it would be difficult to reliably differentiate between a habitat generalist and a species that specialised on the dominant habitat type because both would be found predominantly in the large area habitat. Conversely, for small habitats like the swamp (Figure 4.4 c), it is easier to find positive associations and more difficult to find negative associations because in most random arrangements of individuals, few would be expected in the rare habitat. These points are more prominent for rare species for whom their spatial pattern shows a large variance anyway. In contrast, our method assigns each species to the sub-community it most likely belongs to independent of its abundance and the size of spatial regions. This enables us to explore the habitat preferences of all species, although one should of course be cautious about far reaching conclusions for very rare species that are based on the spatial distribution of only a few individuals. We also note that k-means clustering does force species into different clusters. Whilst we have compared our clustering to that expected under a null model of random associations, our method is more likely to break up large spatial regions due to some species not inhabiting the whole area because of dispersal limitation.

One of the traits revealed by our analysis to strongly influence species distribution at the 50 ha plot at BCI was their shade-tolerance. We found a particularly strong signal for shade-tolerance in juveniles, but in the adult population, shade-tolerance also seemed to be the defining difference between species of the north-western and the south-eastern low plateau sub-communities (Table 4.1). Shade-tolerance is expected to have a stronger impact on the juveniles, because the juvenile population integrates recruitment and survival of individuals over shorter time-scales. By contrast, the adults vary more strongly in age, and because canopy gaps will have opened and closed at different spatial locations over time, adults from different age groups might have been recruited at different spatial locations. Another effect of shade-tolerance is that shade-tolerant and shade-intolerant species generally have very different size-distributions (Wright *et al.*, 2003). While shade-tolerant species maintain large numbers of small non-reproductive individuals in the understorey, individuals of shade-intolerant species either grow to adult stature quickly or die. Importantly, using our method, the weighting of species in the clustering and computation of sub-community densities is independent of each species abundance. This allows us to better compare results from juveniles and adults, as otherwise shade-tolerant species would be dominating the results for the juveniles, while shade-intolerant species would have relatively more weight in the analysis of the adults. This might partly explain why Kanagaraj *et al.* (2011) find weaker effects of the environment on species distribution in adults than in juveniles. Their analysis of the adults is more heavily affected by shade-intolerant pioneer species whose distribution is more strongly linked to canopy gaps than the environmental gradients they included in their study.

In conclusion, we feel there is potentially much benefit in starting with the plants'-eye-views of the landscape and defining spatial regions of interest from clusters of species

with similar spatial associations with one another and with other species. Interpretation of these sub-communities still requires knowledge about both the species and the areas of the landscape they inhabit, and this may or may not be available at the time of study. However, the method can make suggestions for groups of species and areas of the landscape that might merit further attention. As such, we believe our method can increase the understanding of ecosystems that exhibit high biodiversity and for which the spatial habitats that structure the ecosystem and enable the co-existence of such diverse communities are not obvious to the human observer or not, as yet, well understood.

Bibliography

Adler, Peter B, HilleRisLambers, Janneke, & Levine, Jonathan M. 2006. A niche for neutrality. *Ecology Letters*, **10**(2), 95–104.

Bagchi, Robert, Henrys, Peter A, Brown, Patrick E, Burslem, David FR P, Diggle, Peter J, Gunatilleke, CV Savitri, Gunatilleke, IAU Nimal, Kassim, Abdul Rahman, Law, Richard, Noor, Supardi, *et al.* 2011. Spatial patterns reveal negative density dependence and habitat associations in tropical trees. *Ecology*, **92**(9), 1723–1729.

Baldeck, Claire A, Harms, Kyle E, Yavitt, Joseph B, John, Robert, Turner, Benjamin L, Valencia, Renato, Navarrete, Hugo, Davies, Stuart J, Chuyong, George B, Kenfack, David, *et al.* 2013. Soil resources and topography shape local tree community structure in tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, **280**(1753).

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. 2012. Impacts of climate change on the future of biodiversity. *Ecology letters*, **15**, 365–377.

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. 2010. Kernel density estimation via diffusion. *Annals of Statistics*, **38**(5), 2916–2957.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., *et al.* 2002. Habitat loss and extinction in the hotspots of biodiversity. *Conservation biology*, **16**(4), 909–923.

Callaway, Ragan M. 1995. Positive interactions among plants. *The Botanical Review*, **61**(4), 306–349.

Callaway, Ragan M, Brooker, RW, Choler, Philippe, Kikvidze, Zaal, Lortie, Christopher J, Michalet, Richard, Paolini, Leonardo, Pugnaire, Francisco I, Newingham, Beth, Aschehoug, Erik T, *et al.* 2002. Positive interactions among alpine plants increase with stress. *Nature*, **417**(6891), 844–848.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., *et al.* 2012. Biodiversity loss and its impact on humanity. *Nature*, **486**(7401), 59–67.

Comita, Liza S., Muller-Landau, Helene C., Aguilar, Salomon, & Hubbell, Stephen P. 2010. Asymmetric Density Dependence Shapes Species Abundances in a Tropical Tree Community. *Science*, **329**(5989), 330–332.

Condit, Richard. 1998. *Tropical Forest Census Plots*. Berlin, Germany, and Georgetown, Texas: Springer-Verlag and R. G. Landes Company.

Condit, Richard, Ashton, Peter S., Baker, Patrick, Bunyavejchewin, Sarayudh, Gunatilleke, Savithri, Gunatilleke, Nimal, Hubbell, Stephen P., Foster, Robin B., Itoh, Akira, LaFrankie, James V., Lee, Hua Seng, Losos, Elizabeth, Manokaran, N., Sukumar, R., & Yamakura, Takuo. 2000. Spatial Patterns in the Distribution of Tropical Tree Species. *Science*, **288**(5470), 1414–1418.

Condit, Richard, Engelbrecht, Bettina MJ, Pino, Delicia, Pérez, Rolando, & Turner, Benjamin L. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. *Proceedings of the National Academy of Sciences*, **110**(13), 5064–5068.

CTFS. 2013. *Center for Tropical Forest Science*. [online] http://www.ctfs.si.edu. Accessed: 2013-06-15.

Dalling, Jim, John, Robert, Harms, Kyle, Stallard, Robert, & Yavitt, Joe. 2009. *Soil Maps of Barro Colorado Island 50 ha Plot*. `http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/soilmaps/BCIsoil.html` [Online; accessed 06-February-2013].

Flügge, Anton J, Olhede, Sofia C, & Murrell, David J. 2012. The memory of spatial patterns: changes in local abundance and aggregation in a tropical forest. *Ecology*, **93**(7), 1540–1549.

Gan, G., Ma, C., & Wu, J. 2007. Data Clustering: Theory, Algorithms, and Applications (Asa-Siam Series on Statistics and Applied Probability). 2007. *Society for Industrial & Applied Mathematics, USA*.

Haase, Peter. 1995. Spatial Pattern Analysis in Ecology Based on Ripley's K-Function: Introduction and Methods of Edge Correction. *Journal of Vegetation Science*, **6**(4), 575–582.

Hardin, Garrett. 1960. The Competitive Exclusion Principle. *Science*, **131**(3409), 1292–1297.

Harms, Kyle E., Condit, Richard, Hubbell, Stephen P., & Foster, Robin B. 2001. Habitat Associations of Trees and Shrubs in a 50-Ha Neotropical Forest Plot. *Journal of Ecology*, **89**(6), 947–959.

Hubbell, Stephen P. 2001. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. Princeton University Press.

Hubbell, Stephen P., Foster, Robin B., O'Brien, S. T., Harms, K. E., Condit, Richard, Wechsler, B., Wright, S. J., & de Lao, S. Loo. 1999. Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science*, **283**(5401), 554–557.

Hubbell, Stephen P., Condit, Richard, & Foster, Robin B. 2005. *Barro Colorado Forest Census Plot Data*. [online] https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Isbell, F.I., Polley, H.W., & Wilsey, B.J. 2009. Biodiversity, productivity and the temporal stability of productivity: patterns and processes. *Ecology letters*, **12**(5), 443–451.

Itoh, Akira, Ohkubo, Tatsuhiro, Nanami, Satoshi, Tan, Sylvester, & Yamakura, Takuo. 2010. Comparison of statistical tests for habitat associations in tropical forests: A case study of sympatric dipterocarp trees in a Bornean forest. *Forest Ecology and Management*, **259**(3), 323–332.

Kanagaraj, Rajapandian, Wiegand, Thorsten, Comita, Liza S., & Huth, Andreas. 2011. Tropical tree species assemblages in topographical habitats change in time and with life stage. *Journal of Ecology*, **99**(6), 1441–1452.

Law, Richard, Illian, Janine, Burslem, David FRP, Gratzer, Georg, Gunatilleke, CVS, & Gunatilleke, IAUN. 2009. Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology*, **97**(4), 616–628.

Ledo, Alicia, Burslem, David FRP, Condés, Sonia, & Montes, Fernando. 2013. Micro-scale habitat associations of woody plants in a neotropical cloud forest. *Journal of Vegetation Science*, **24**(6), 1086–1097.

Muller-Landau, Helene C., & Hardesty, Britta D. 2005. Seed dispersal of woody plants in tropical forests: concepts, examples and future directions. *Pages 267–309 of:* Burslem, Pinard, & Hartley (eds), *Biotic interactions in the Tropics*. Cambridge: Cambridge University Press.

Murrell, David J., Purves, Drew W., & Law, Richard. 2001. Uniting pattern and process in plant ecology. *Trends in Ecology & Evolution*, **16**(10), 529–530.

Peters, H.A. 2003. Neighbour-regulated mortality: the influence of positive and negative density dependence on tree populations in species-rich tropical forests. *Ecology Letters*, **6**(8), 757–765.

Punchi-Manage, Ruwan, Getzin, Stephan, Wiegand, Thorsten, Kanagaraj, Rajapandian, Savitri Gunatilleke, CV, Nimal Gunatilleke, IAU, Wiegand, Kerstin, & Huth, Andreas. 2013. Effects of topography on structuring local species assemblages in a Sri Lankan mixed dipterocarp forest. *Journal of Ecology*, **101**(1), 149–160.

Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(30), 10854–10861.

Uriarte, M., Condit, Richard, Canham, C. D., & Hubbell, Stephen P. 2004. A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? *Journal of Ecology*, **92**, 348–360.

Valencia, R., Foster, R.B., Villa, G., Condit, R., Svenning, J.C., Hernández, C., Romoleroux, K., Losos, E., Magård, E., & Balslev, H. 2004. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *Journal of Ecology*, **92**(2), 214–229.

Volkov, I., Banavar, J.R., Hubbell, S.P., & Maritan, A. 2009. Inferring species interactions in tropical forests. *Proceedings of the National Academy of Sciences*, **106**(33), 13854–13859.

Wiegand, T., & Moloney, K.A. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**(2), 209–229.

Wiegand, Thorsten, Huth, Andreas, Getzin, Stephan, Wang, Xugao, Hao, Zhanqing, Gunatilleke, CV Savitri, & Gunatilleke, IAU Nimal. 2012. Testing the independent species arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1741), 3312–3320.

Williams, J.N. 2012. Humans and biodiversity: population and demographic trends in the hotspots. *Population & Environment*, **[in press]**.

Wright, J.S. 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, **130**(1), 1–14.

Wright, S Joseph, Muller-Landau, Helene C, Condit, Richard, & Hubbell, Stephen P. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology*, **84**(12), 3174–3185.

Wright, S Joseph, Kitajima, Kaoru, Kraft, Nathan JB, Reich, Peter B, Wright, Ian J, Bunker, Daniel E, Condit, Richard, Dalling, James W, Davies, Stuart J, Díaz, Sandra, *et al.* 2010. Functional traits and the growth-mortality trade-off in tropical trees. *Ecology*, **91**(12), 3664–3674.

Chapter 5

Structure and stability of the normalised co-associations matrix

Summary

The spatial distribution of tree species in a forest holds information on the environment and ecological processes that structure the ecosystem. While past attempts to study the spatial patterns of species have often focused on individual species or pairs of species we are interested in processes that shape the spatial associations between whole sub-communities of species in the ecosystem. In the previous chapter, we introduced a normalised co-association matrix that contains information on the aggregation or segregation between spatial patterns of each pair of species. In that chapter we focused on analysing a single normalised co-association matrix and deriving clusters of species and sub-community maps of how those clusters of species are spatially distributed in the forest. However, when multiple data sets from different census time points, individuals of different life-stages, different sets of species, or even data from different forests are available, we can compute multiple co-association matrices. Here we develop further methods to analyse and visualise the structure of those matrices and to allow between-matrix comparisons. We find that shade-intolerance causes strong spatial co-aggregation, but only in juveniles and recruits and not in adults, and that habitat dependency leads to spatial segregation of species in juveniles and adults. We do not find any indication that phylogenetic relatedness has a net impact on the normalised co-association values between species pairs. The co-association matrices of juveniles and adults from two censuses and the derived clusters of species and sub-community maps become less similar if the censuses are more distant in time. This is not true for co-association matrices from recruits, which show very little stability of sub-community maps, but for which species clusters are markedly more stable and independent of the time between compared data sets.

Impact of this work

The work presented in this chapter builds on the methods introduced in the previous chapter and shows various ways to further analyse the normalised co-association matrix. In particular we introduce tools to compare multiple co-association matrices, and the results that build on it, based on different data sets. This is important because only by making comparisons between different ecological data sets, we will be able to identify the general principles that govern those systems. As more and more high quality spatial point pattern data sets from many different forests and other ecosystems become available we are now in the position to make such comparisons. The work presented here can contribute a valuable tool in that pursuit.

Declaration on the contributions to the work presented in this chapter:

All analyses were conducted by me. The chapter was written by me. Sofia Olhede and David Murrell contributed by supervising and guiding my work and by providing feedback on a draft of this chapter. The data used was provided by the Centre for Tropical Forest Science.

5.1 Introduction

In the previous chapter we have introduced a novel method to group species into subcommunities depending on their spatial co-associations. In that chapter our focus was on finding groups of species and on investigating the dominance of the various groups at different places in the study area. This analysis was based on a matrix of normalised bi-variate co-associations between all pairs of species (defined in Equation 4.2). However, the structure of this normalised co-associations matrix itself is of interest because it might reveal structural properties of the ecosystem we are studying, such as the importance of gap dynamics or habitat filtering. In this chapter we study this matrix in more depth and present various methods to visualise its structure and stability. One important feature of the normalised co-association matrix is the normalisation procedure we use to make the co-association values of different species pairs comparable. For the normalisation we use a null-model that assumes that the species' spatial patterns follow independent inhomogeneous Poisson distributions. We use random torus translations of the species' spatial patterns relative to each other to estimate and correct for the expected variance in between-species co-association under the null-model, while keeping the marginal characteristics of the within-species spatial pattern constant. This is necessary to compare co-association values of species with different abundances and within-species aggregations within the same matrix, but it also allows us to compare co-association values across matrices derived from different spatial point pat-

tern data sets. Even more generally, it allows us to compare whole co-association matrices with each other. Most methods presented in this chapter can be used to analyse and visualise the structure of a single normalised co-association matrix, but in doing so, they also suggest themselves to be used to compare multiple co-association matrices and this possibility will be explored in this chapter.

In the first part of this chapter we examine normalised co-association matrices and factors that may shape them. Among the factors that are thought to influence the spatial co-association of species are shade-tolerance, life-stage, habitat dependency and phylogenetic relatedness. For each of those factors we investigate how they affect the structure of the co-association matrix. First, shade-intolerance is thought to cause strong co-aggregation between species in light gaps during recruitment (Schnitzer & Carson, 2001). However, co-aggregation between shade-intolerant species is expected to decrease as individuals progress through life-stages and canopy gaps close (see Chapter 4). We therefore predict that shade-intolerant species will show stronger positive co-associations with other species as recruits and juveniles than as adults and we will test if this relationship can be used to predict the shade-intolerance of species. Second, habitat dependency should cause both co-aggregation of species with similar habitat preferences as well as segregation of species with different habitat requirements. This effect of habitat on species pattern might increase with the age of individuals because habitat filtering works over time, although the evidence for this at BCI is mixed; Kanagaraj *et al.* (2011) found the spatial distribution of adults to be more homogeneous and less dependent on environmental variables than the spatial distribution of juveniles, and Comita *et al.* (2007) showed that for some species, habitat preferences change between life-stages. To explore this relationship further, we will compare co-association values of species known to be more habitat dependent and species less habitat dependent (Harms *et al.*, 2001), separately for recruits, juveniles and adults. The strength of both positive and negative co-associations could also be used to investigate how much of an effect habitat dependency generally has in a specific ecosystem. We compare the normalised co-association matrix from the BCI data to a random null-model without any habitat structure to see how much stronger associations are at BCI. While we only use this approach to illustrate how different BCI is from that one model, the method could also be used to compare the data to a variety of different models and select a best fitting model. For example we could compare models with a different gap dynamic and investigate how the number and size of light gaps in the canopy may affect the co-association matrix and we could see which parameters on the gap dynamics fit the empirical data best. Furthermore, this approach would also allow us to compare different forests and thereby test hypotheses on the strength of various processes at different sites. Finally, phylogenetic relatedness might also cause both segregation and aggregation (see Swenson *et al.*, 2007; Pearse *et al.*, 2013; Mayfield & Levine, 2010). When biological traits are conserved in closely related species they are more likely to be fierce

competitors and share natural enemies, which could lead to segregation. But they may also share the same preferences for environmental conditions and therefore show similar habitat dependencies. It is not clear whether on balance we would expect competitive exclusion to cause segregation or shared preferences and similar competitive abilities to cause aggregation between close relatives (Mayfield & Levine, 2010). By contrasting phylogenetic relatedness with co-association we test if we can find evidence for either.

The analyses in the second part of this chapter examine the stability of the derived clusters of species (henceforth referred to as "sub-communities") and the sub-community maps derived from multiple co-association matrices. To this end we will look at co-association matrices from different time-points and based on individuals from different life-stages. While in most cases it might be reassuring if a result is stable across time and life-stages, there are also cases where we expect changes or where changes can be informative about processes or events that shape the structure of the ecosystem. For example we might expect that large disturbances such as a strong drought caused by a strong El Niño year have a substantial impact on the spatial co-associations between species (although evidence for the longer-term impact of droughts at BCI is limited; see Leigh Jr *et al.*, 1990; Condit *et al.*, 2004). Another important question is whether a directional change in the environmental conditions over time (Feeley *et al.*, 2011) will alter spatial co-associations between species over time. Depending on the process that causes a particular co-association pattern we might also expect differences between life-stages. Sub-communities that are shaped by shared habitat requirements may become more pronounced from recruits to juveniles and adults, while shade-intolerant gap-dependent species might be most strongly co-associated in the recruit stage. For habitat dependent sub-communities the spatial area which they dominate might be fairly constant across time, for gap-dependent species on the other hand, the spatial distribution might show rapid changes while the species composition of sub-communities might be far more stable.

All results in this chapter are based on the data from the 50 ha Barro Colorado Island forest dynamic plot in Panama (Hubbell *et al.*, 2005). However, the potential of the methods presented here will only be fully utilised when compared with data from other ecosystems. Such extensions of the methods presented here for cross-ecosystem comparisons are discussed.

5.2 Methods

5.2.1 Data used from Barro Colorado Island (BCI)

In this chapter we use the data from the 50 ha BCI forest dynamic plot (Hubbell *et al.*, 2005) to compute the normalised co-association matrix as defined in Equation 4.2 in Chapter 4 which makes co-association values of pairs of species with different abundances and within-species spatial patterns comparable. We use the data on the spatial distribution of all individuals above 1 cm diameter at breast height (DBH) for the 141 tree and shrub species that are

present within the plot with at least 10 juveniles and 10 adults in the 2010 census. Individuals are classified as juveniles or adults using a species specific DBH threshold by Robin Foster estimating the minimum DBH at the time of first flowering. Depending on the analysis, we restrict the set of species further to those species for which additional information on shade-tolerance, habitat dependency or phylogenetic relatedness is available, and for which there is also a minimum of at least 10 individuals available at different censuses or at the recruit stage. Individuals are classified as recruits in a specific census if they first appear in that census (i.e., reach a DBH of more than 1 cm) and have not been present in any previous census (Kanagaraj *et al.*, 2011). It should be noted that the time between censuses is five years in all but the first inter-census period, which was only 2 years.

5.2.2 Distribution of values in the co-association matrix

Having computed the normalised co-association matrix as described in Chapter 4 we can use different methods to visualise and further analyse it. In this chapter we will present four ways to explore the relationship between the normalised co-associations matrix and properties of the species that might shape their co-association structure.

5.2.2.1 Sorting the normalised co-association matrix

One of the simplest ways of visualising structure in the co-association matrix is by sorting its rows and columns according to a variable whose effect on the spatial structure we want to investigate. This can be done by grouping species in the same categories together (see Figure 4.1) or by ordering them according to an ordinal variable. This visualisation is effective to quickly examine the evidence for an effect of an individual variable on the structure in the co-association matrix. Here we present an analysis on how shade-tolerance influences the co-associations of species in adults and juveniles. We compute the normalised co-association matrix for all 124 species in the 2010 census for which information on the shade-tolerance was available. For an estimate of the shade-tolerance of species we use the shade-tolerance index reported by Comita *et al.* (2010). This measure is based on the growth and mortality rates of 1 to 5 cm saplings in the BCI plot between 2000 and 2005 and assigns each species the value of the first axis in a principle component analysis of the growth-mortality space. As shade-intolerant species are expected to put more resources in growth and less in survival the combination of growth and death rate is a good indicator for shade-tolerance. We use this shade-tolerance index to sort the rows and columns of the normalised co-association matrices that we computed independently for adults and juveniles.

5.2.2.2 Sum of differences between normalised co-association matrices

Based on the same two normalised co-association matrices mentioned in the previous section, we compute the sum of differences between the co-association of the juveniles and the adults

P_a for each species a. P_a being defined as:

$$P_a = \sum_{b \in \text{Species}} \left(\bar{\Omega}^{\text{Juveniles}(a,b)}_{0,10,2010} - \bar{\Omega}^{\text{Adults}(a,b)}_{0,10,2010} \right) \tag{5.1}$$

Here $\bar{\Omega}^{\text{Juveniles}(a,b)}_{0,10,2010}$ is the normalised co-association value between the juveniles of species a and b, and $\bar{\Omega}^{\text{Adults}(a,b)}_{0,10,2010}$ between the adults of said species (note that the within-species normalised co-association value is defined to be zero in Equation 4.2; changes in the within-species pattern between life-stages therefore do not contribute to P).

Having computed P_a for each of the 124 species for which the shade-tolerance index (Comita *et al.*, 2010) is available, we contrast the two measures to examine if P_a can be used as a predictor for shade-tolerance. To this end we use a linear regression model to assess how much of the variance in the shade-tolerance index ST can be predicted by P:

$$\boldsymbol{ST} = \beta_0 + \beta_1 \cdot \boldsymbol{P} + \boldsymbol{\epsilon}. \tag{5.2}$$

With $\boldsymbol{ST}$ being the vector of shade-tolerance indices ST_a for all species a in the analysis, $\boldsymbol{P}$ the vector of values P_a for all species a, and $\boldsymbol{\epsilon}$ a vector of error terms with zero mean which are assumed independent between species. We use a bootstrapping method as described in Section 3.2.4 to compute if $\boldsymbol{P}$ can explain a significant amount of variance in $\boldsymbol{ST}$.

5.2.2.3 Relative density distribution of normalised co-association values

A way to summarise the two dimensional matrix of normalised co-association values in one dimension is to plot the relative density distribution of values in the matrix, i.e., to compute the probability of an entry in the matrix having a certain value. We use Botev's kernel density estimator (Botev *et al.*, 2010) to estimate the probability density curve from the normalised co-association values. This visualisation can be used to compare the properties of different co-associations matrices with each other. Here we use this method to compare the co-association values for adults and juveniles in the 2010 census and contrast them to a torus-translation null-model. For the torus-translation model we compute 1000 random forests for which we use an independent uniformly random torus-translation to shift the spatial pattern of all species relative to the pattern of the other species. We then compute the co-association matrix and the probability density function for all random forests and report the 2.5 and 97.5 percentiles of their probability functions at any position.

In a second analysis we compare the effect of habitat dependence between adults, juveniles and recruits. For this analysis we group species in habitat dependent and habitat independent species depending on whether they were found to be significantly correlated (positively or negatively) in the torus translation test by Harms *et al.* (2001) with at least one habitat type (we excluded species that were not included in Harms *et al.*, 2001). We use the data of the 110 species with more than 10 adults, 10 juveniles, and 10 recruits in the 2010

census for which information on habitat dependency was available. 73 of those species were classified as habitat dependent and 37 were classified as habitat independent. We then independently compute the normalised co-association matrices for habitat dependent and habitat independent species at each of the three life-stages.

5.2.2.4 Properties of species pairs

If we are interested in whether a property of pairs of species might explain their co-association values, we can plot a scatter plot of species pairs with the value of the property of interest on one axis and the normalised co-association value of the pair on the second axis. In Figure 5.5 we use this visualisation to inspect if there is a relationship between the phylogenetic distance of two species and their spatial co-association. We use the phylogenetic data from Kress *et al.* (2009) derived from DNA analysis. Based on their phylogenetic tree we compute the distance between each pair of species and plot this distance against the normalised co-association values of the adults of the selected species in the 2010 census. We include all 131 species which were included in the phylogeny by Kress *et al.* (2009).

5.2.3 Stability of sub-communities and sub-community maps

In our final analysis we explore the stability of the assignment of species to sub-communities using k-means clustering, and the consistency of the maps of dominant sub-communities. For this analysis we compute the normalised co-association matrices independently for recruits, juveniles, and adults in the six censuses in 1985, 1990, 1995, 2000, 2005, and 2010, for all 93 species with at least 10 recruits, 10 juveniles, and 10 adults in each of the censuses.

Having computed the co-association matrices we then group species into sub-communities using k-means clustering for cluster numbers between $k = 2$ and $k = 9$ as described in Section 4.2.2.3. In a further step we also compute the dominant sub-community maps as described in Section 4.2.2.4 and Section 4.2.2.5. Based on the obtained clustering and the resulting sub-community maps, we then determine their stability over time or between life-stages by comparing them between censuses and between adults, juveniles, and recruits. First, we match each sub-community found in the 2010 census with the best matching sub-community at the previous censuses by computing which cluster at each time-point has the maximum spatial overlap to which cluster at the other time-point. We can then calculate the number of 20x20 m quadrants H_M that are classified as the best matching cluster at the different census and thereby obtain a measure of the stability of the clustering over time. Because we would always expect to find some overlap if we match the best matching clusters of any two sub-community maps, we correct the raw size of the overlap by subtracting the mean overlap H_R. H_R is obtained by doing multiple torus translations of one of the sub-community maps and computing the overlap with the translated map. The resulting corrected measure of overlap is then divided by the number of 20x20 m quadrants in the plot

N_Q minus the mean overlap, i.e., we compute

$$C_H = \frac{H_M - H_R}{N_Q - H_R} \tag{5.3}$$

Thereby, we obtain a measure for the percentage of overlap that is larger than expected for unrelated random maps of that structure. This means for unrelated maps, the expected value of C_H would be zero and for identical maps C_H would be equal to one. In this study we used the mean of all possible 1249 unique torus translations (translating the map in multiples of 20 m in the east-west and the north-south direction) to compute H_R.

Similar to the described measure of spatial overlap of dominant clusters in sub-community maps we can define a measure of species overlap C_S between sub-communities at different time-points. Instead of counting the number of 20x20 m quadrants H_M that are dominated by the best matching cluster at another time-point, we can count the number of species S_M that are shared between sub-communities at different time-points. Instead of the torus translation in the case of the habitat maps we used 1000 random permutations of the associations between the species and the clusters (keeping the number of species in each cluster constant) to compute the expected overlap S_R under a null-model of no association. C_S is accordingly defined as

$$C_S = \frac{S_M - S_R}{N_S - S_R} \tag{5.4}$$

with N_S being the number of species.

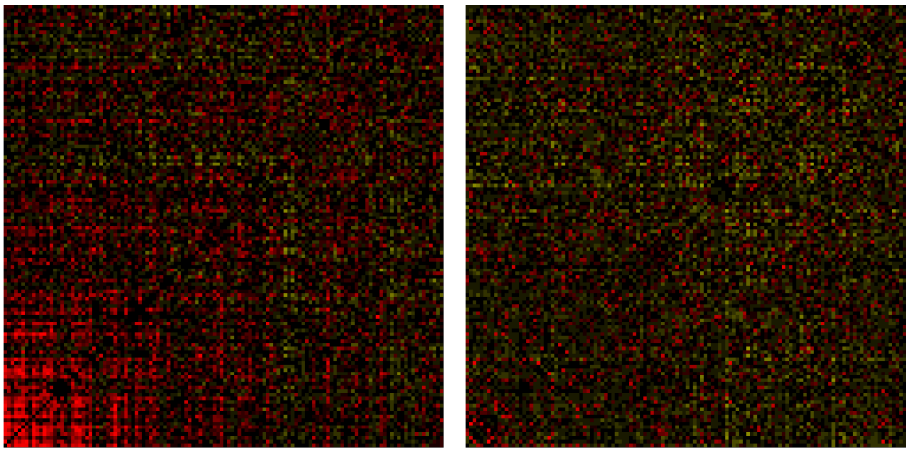
The above computation was repeated independently for adults, juveniles, and recruits, in each case comparing the 2010 census with the 5 censuses before in 1985, 1990, 1995, 2000, and 2005.

5.3 Results

5.3.1 Distribution of values in the co-association matrix

5.3.1.1 Sorting the normalised co-association matrix

In Figure 4.1 we have sorted the rows and columns of the co-association matrix with respect to the clustering found by k-means algorithm. This allowed us to immediately see that within-cluster co-aggregation is generally much larger than between-cluster co-aggregation; thereby confirming that our method can indeed identify sets of species that are co-aggregated. By sorting the rows and columns according to different criteria we can highlight different structures. In Figure 5.1 we display the co-association matrix for adults and juveniles sorted by the shade-tolerance of the species. It can be seen that shade-intolerant species are particularly co-aggregated as juveniles, but that this effect diminishes in the transition to the adult life-stage. This result supports the finding in Chapter 4 where a cluster of shade-intolerant pioneering species was consistently identified in the juveniles, but not in the adults.



(a) Juveniles (b) Adults

Figure 5.1: Normalised co-association $\bar{\Omega}^{i,j}_{0,10,2010}$ matrix between all pairs of the 124 species at Barro Colorado Island (BCI) with more than 10 juveniles and 10 adults in the 2010 census and for which shade tolerance estimates are available. The matrix is sorted by shade tolerance (from the most shade-intolerant at the bottom-left to the most shade-tolerant at the top-right). Red indicates positive values (co-aggregation), yellow negative values (segregation), black neither aggregation nor segregation.

5.3.1.2 Sum of differences between normalised co-association matrices

As observed in Figure 5.1, shade-tolerance leads to much higher co-aggregation between the juveniles of shade-intolerant species than it does in the adults. In Figure 5.2 we quantify the relationship between shade-tolerance and differences in co-associations of juveniles and adults. It can be seen that generally, the more shade-intolerant species show a larger difference between juvenile and adult co-associations. The sum of the differences between the co-associations $\boldsymbol{P}$ can explain 40.1% of the variance in the shade-tolerance index (R^2). This result is highly significant; the percentage of variance explained is higher than in any of 10000 randomised bootstrap samples we computed to estimate the significance level.

5.3.1.3 Relative density distribution of normalised co-associations values

In Figure 5.3 we show the distribution of normalised co-association values for adults compared to juveniles. This demonstrates that juveniles show stronger positive co-associations than adults. From Figure 5.1 we know that those high co-aggregations occur mainly between shade-intolerant species, i.e., new individuals pre-dominantly recruiting in light gaps. The results for the adults suggest that through thinning during succession, when a canopy gap closes, this pattern largely disappears. In adults there is only a weak effect of additional co-aggregation between species compared to the results from the random forests null-model. Both juveniles and adults show more than expected segregation between species when compared to the null-model. This might be caused by direct competition between species, but could also be due to habitat dependency and therefore spatial segregation of species specialised on different habitats. The latter hypothesis is supported by Figure 5.4 that shows no systematic difference between habitat dependent and habitat independent species in the

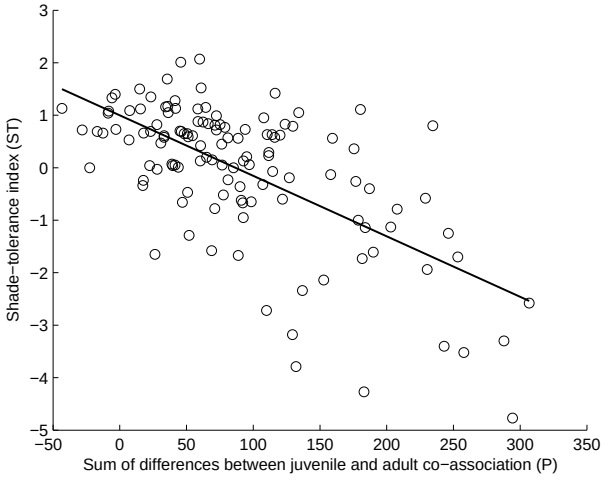


Figure 5.2: Each circle represents one species in the 2010 census. Depicted is the relationship between the sum of the differences P_a between the co-association values of species a in the juveniles with that value in the adult population and the shade-tolerance index as defined by Comita *et al.* (2010). The line shows the linear least square fit to the data.

number of strong positive co-associations, but finds that species that are habitat dependent are more often segregated from other habitat dependent species in both the juvenile and adult populations than non habitat dependent species are among themselves.

5.3.1.4 Properties of species pairs

In Figure 5.5 we present another way to analyse the normalised co-association matrix with respect to other available information, in particular with respect to another variable about the difference or relationship between two species. Figure 5.5 depicts the phylogenetic difference between two species with respect to their normalised co-association value. There is no apparent relationship between the two factors, i.e., closely related species neither seem to be particularly co-aggregated nor segregated in space.

5.3.2 Stability of sub-communities and sub-community maps

Figure 5.6 and Figure 5.7 show how stable the grouping of species and the resulting derivation of sub-community maps is through time. As expected we can confirm that if less time passed between two censuses the results are generally more similar. Figure 5.6 looking at the sub-community maps shows that those maps are more consistent in the adults than the juveniles, and much more consistent in the adults than in the recruits. This should not be a surprise to us as generally the adult population is changing less between censuses; but it could also point to processes of habitat filtering whereby species become more constrained to their preferred habitat during the progression through life-stages. However, if we look at the consistency of the species groups in Figure 5.7, the difference between recruits and adults is much smaller and the stability for the juveniles is generally as strong if not stronger than for the adults. This suggests that habitat filtering after the recruit stage might not explain most of the difference between adults and recruits in the sub-community maps because we would expect habitat

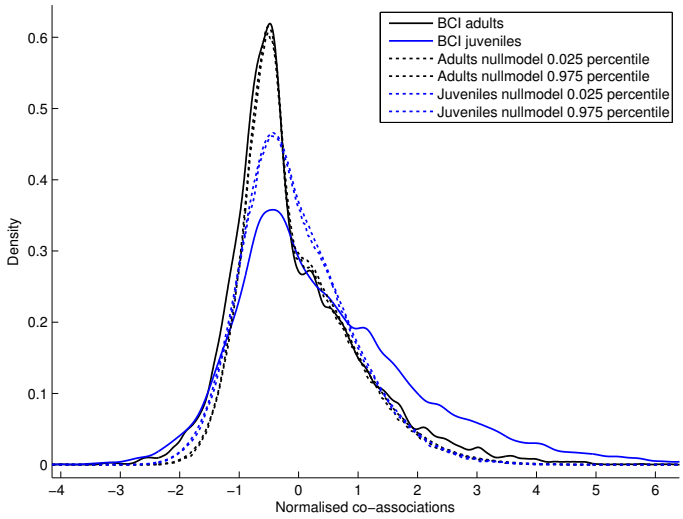


Figure 5.3: Distribution of the values of the upper triangle of the normalized cross $\bar{\Omega}^{i,j}_{0,10,2010}$ matrix without the diagonal for the 141 species at BCI forest dynamics plot with more than 10 adults and 10 juveniles in the 2010 census. The solid black line shows the result for the adult individuals, the solid blue line for the juveniles. Dotted lines show the 2.5 and the 97.5 percentiles for 1000 random forests realisations obtained by random torus-translation shifts of the spatial pattern of each species in the analysis.

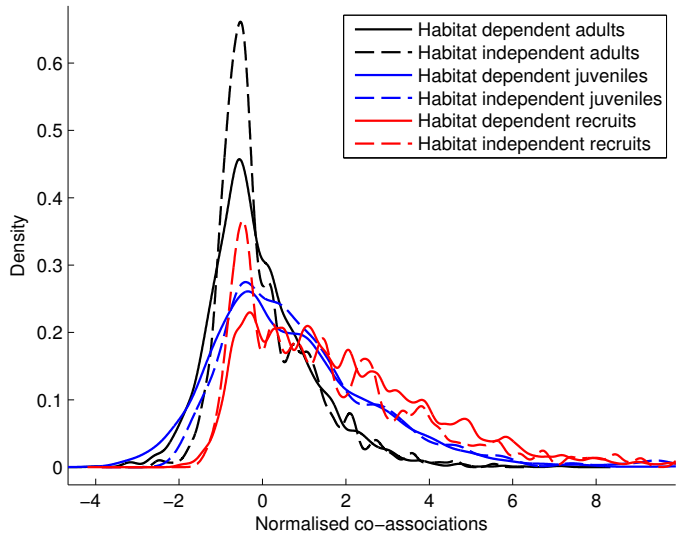


Figure 5.4: Distribution of the values of the upper triangle of the normalized cross $\bar{\Omega}^{i,j}_{0,10,2010}$ matrix without the diagonal between all pairs of species with more than 10 adults, 10 juveniles and 10 recruits in the 2010 census. The 37 species that were not found to be habitat dependent by Harms *et al.* (2001) are depicted by dashed lines and the 73 species which were found to be habitat dependent by solid lines. The results for adults are shown in black, the results for the juveniles in blue, and the results for the recruits in red.

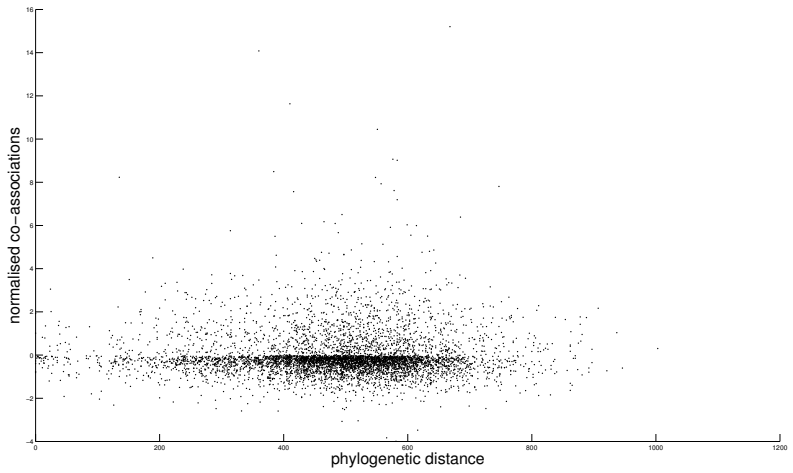


Figure 5.5: Depicted are the normalized cross $\bar{\Omega}^{i,j}_{0,10,2010}$ values for the adults versus the phylogenetic distance (using the phylogeny of Kress *et al.* (2009)) between all pairs of species of those 131 species with more than 10 adults and 10 recruits in the 2010 census which were included in the phylogeny.

filtering to affect the results for the species clustering in a similar way. Instead the result might be explained by changing environmental conditions over time that are averaged out in the adult population, but not in a single recruit generation. Depending on canopy gaps and other changing environmental conditions, such as water availability, different recruitment cohorts might prefer different spatial regions; but the same species still find themselves together in what are favourable conditions for them in each generation. This could explain why the recruit clusters are relatively stable in their species composition but rapidly changing in the spatial regions they dominate.

Further we notice that for larger numbers of clusters k the results seem to become more unstable for the sub-community maps (see Figure 5.6) but remain relatively unchanged for the species composition (see Figure 5.7). This could be due to sub-communities of species overlapping in their spatial extent. The sub-community map only contains the dominating cluster at any point in space. This measure is meaningful and expected to be relatively stable as long as sub-communities are sufficiently distinct in space. If multiple sub-communities overlap and are nearly equally represented in a particular region, chance events become more important in determining which sub-community is deemed dominant in that region and the assignment is more likely to change between time points. Obviously, if sub-communities strongly overlap they might at some point not be distinguishable at all anymore and the stability of the species assignments to clusters should suffer as well. However, Figure 5.7 suggests that there are sub-communities that, despite being partially overlapping in space, are still sufficiently distinct in their species composition to be detected even for numbers of clusters at which the stability of the sub-community maps is deteriorating.

It should be noted that even though we can qualitatively compare relative differences in stability between different life-stages and numbers of clusters for the sub-community maps

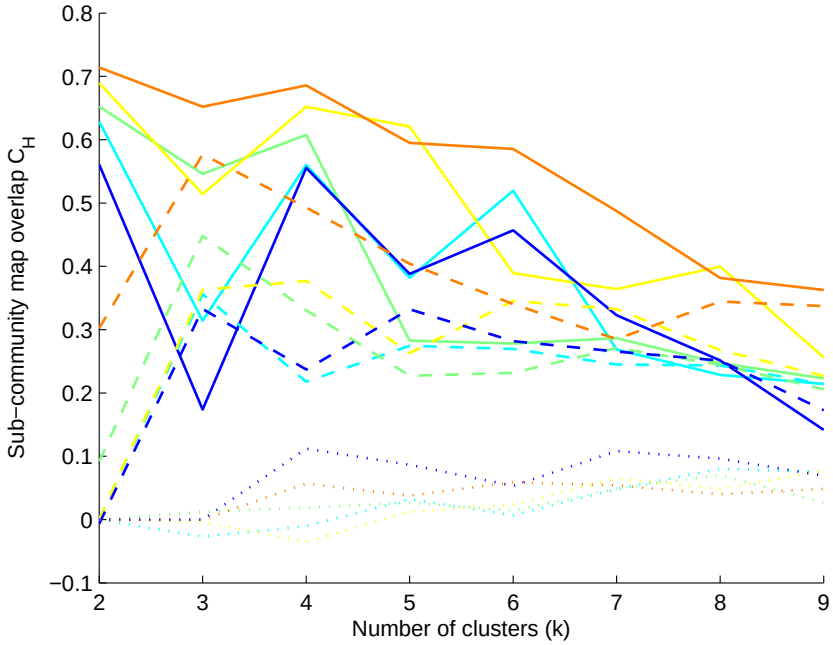


Figure 5.6: The spatial stability of the sub-communities between 1985 and 2010. Depicted are the consistencies between the sub-community map derived from the 2010 census data and earlier time points as explained in the methods. Solid lines show results for adults, dashed for juveniles, and dotted for recruits. The different colours show different censuses which we compared to the 2010 census: 1985 (dark blue), 1990 (light blue), 1995 (green), 2000 (yellow), 2005 (orange).

and the species composition, we should not directly compare the absolute values for the stability between the two measures. The two measures are more different than it might seem at first. The sub-community maps have a spatial structure of neighbouring locations that we take into account by using a torus-translation null-model. By contrast, for species composition, we only consider the distribution of species numbers in the different clusters, but we do not consider any structure in the species relationships.

5.4 Discussion

In this chapter we have presented various ways to visualise and analyse the normalised co-association matrix we introduced in the previous chapter. We plotted the sorted matrix in Figure 5.1 to show how the structural impact of a property of the species can be illustrated. With this visualisation we could highlight the difference between juveniles and adults and we could show that while shade-tolerance was a dominant factor in shaping the co-association of species in juveniles, this effect was dramatically diminished in the adult population. The finding that shade-intolerant species co-aggregate in light gaps in early stages of their life is expected and also that this effect diminishes over time, but the extent to which the co-aggregation is reduced is surprising. Indeed, the magnitude of the change of co-association values between juveniles and adults could be used as a proxy for shade-tolerance as 41% of

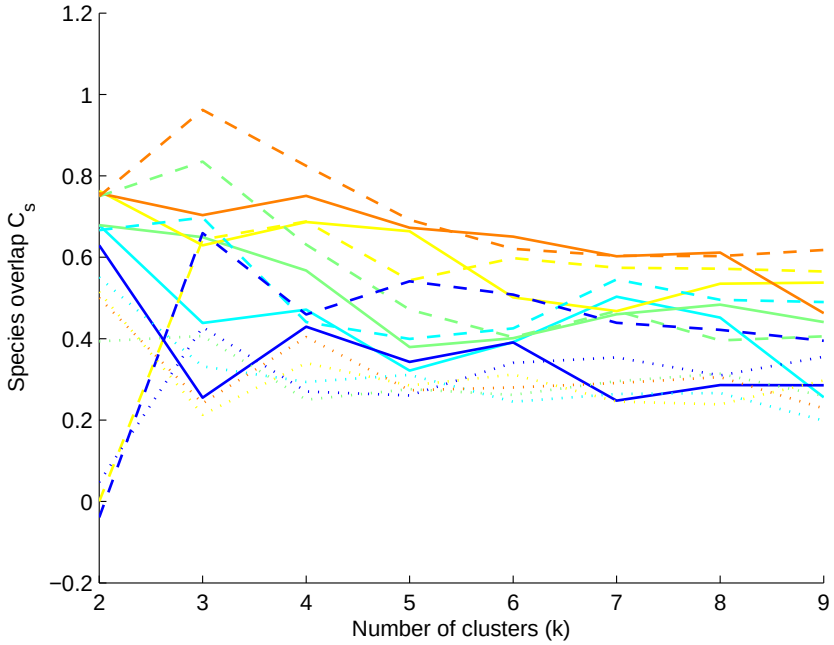


Figure 5.7: The group membership stability of the sub-communities between 1985 and 2010. Depicted are the consistencies between species composition of sub-communites derived from the 2010 census data and earlier time points as explained in the methods. Solid lines show results for adults, dashed lines for juveniles, and dotted lines for recruits. The different colours show different censuses which I compared to the 2010 census: 1985 (dark blue), 1990 (light blue), 1995 (green), 2000 (yellow), 2005 (orange).

the variance in the shade-tolerance index could be explained by the sum of the differences between the juvenile and adult normalised co-association values of a species (see Figure 5.2). We note that the shade-tolerance index as defined by Comita *et al.* (2010) is computed from the growth and survival rates of saplings between the 2000 and 2005 censuses within the BCI forest dynamic plot. As spatial co-associations might affect growth and survival, the estimation of the shade-tolerance index is not entirely independent of the spatial pattern observed in the 2010 census. However, interspecific species competition at BCI is weak (Comita *et al.*, 2010; Wiegand *et al.*, 2012) and therefore even highly co-aggregated species are not expected to strongly bias the estimation of the shade-tolerance index. Other cases in which sorted normalised co-aggregation matrices could be used as a visualisation tool could be when exploring the influence of other species traits, such as tolerance to certain soil variables (Condit *et al.*, 2013), or analogous to the comparison of juveniles and adults presented here, the comparison of different forests.

By using probability density plots of the normalised co-association matrices as in Figure 5.3 and Figure 5.4 we lose the information on the structure presented in Figure 5.1, i.e., on which co-association values belong to the same species. However, we gain the ability to contrast different communities or sub-communities within one graph. This allows us to compare the distribution of co-associations between different life-stages and between empirical

data and model results as in Figure 5.3, as well as between different sub-groups of species as in Figure 5.4. In principle, this visualisation can be used to compare any number of different normalised co-association matrices based on different data-sets, models, or species selection criteria. The analysis of the data shown here confirms that a higher than expected co-aggregation between species compared to a null-model of no association is mainly found in juveniles, while segregation is found both in juveniles and adults. Figure 5.4 shows that segregation in both juveniles and adults is more common between habitat dependent species than between habitat independent species, but not more common in recruits. This finding could be interpreted as confirmation of previous results by Kanagaraj *et al.* (2011) showing that habitat filtering at BCI operates mainly at the transition from the recruit to the juvenile stage.

A third way to look at the values in the normalised co-association matrix is to plot the co-association values against a property of the species pair. Because the co-association value itself is a property of the species pair, rather than an individual species, it suggests itself to investigate if other properties of the pair relationship can explain some of the co-association. Any measure of difference or commonality between the two species in a pair could potentially be presented in this way. In Figure 5.5 we have exemplarily shown this visualisation method with the phylogenetic distances of the species. There does not seem to be any systematic difference in co-associations between closely related species and more distant relatives. However, we should note that this does not prove that phylogeny does not influence spatial co-associations. It could also mean that competing factors cancel each other out. Previous research (Swenson *et al.*, 2007; Pearse *et al.*, 2013) has shown that at small spatial scales there is evidence of phylogenetic over-dispersion caused by more severe competition or shared pathogens of close relatives. By contrast, at larger scales phylogenetic clustering occurs because conserved functional traits lead to preference for similar habitats.

In the second part of this chapter we have explored ways to analyse the stability of grouping species into sub-communities and the stability of sub-community maps. This can be useful both to judge the reliability of results we do not expect to change between co-association matrices based on different data sets, and to investigate how expected or observed changes in the data influence the results from the analysis of the co-association matrices. In this work, we have explored the stability of the clustering and sub-community maps over time, and we have compared that stability between sub-populations of different life-stages. We could also have turned this procedure around and first compared the stability of the clustering and sub-community maps between life-stages and then compared the results for different time-points. This could for example be done to try to answer questions such as how recruitment is affected by external events like a strong El Niño year or the construction of a road nearby.

The analysis presented here confirms some expected results. First, consistency between time-points generally decreases when more time has passed between measurements. Sec-

ondly, the sub-community maps derived from the spatial pattern of adults (in which many individuals will be identical between the time-points in our analysis), are more stable than those for juveniles or recruits. The near total lack of stability of the sub-community maps for recruits is surprising. The definition of a recruit implies that there will be no overlap between different recruit cohorts and that each map is derived from a completely different set of individuals. Nevertheless, one might expect that the spatial pattern of the recruits would be linked more closely to the more stable pattern of seed-producing adults and therefore be more stable in itself. However, local seed-dispersal of the adult population (Muller-Landau & Hardesty, 2005) does not seem to be a sufficient restriction for stable recruit maps at this scale of analysis. As supported by Figure 5.4 and Kanagaraj *et al.* (2011), habitat filtering also does not seem to enforce a consistent spatial structure in the recruit stage. However, this does not necessarily imply that habitat filtering is not important in the establishment of individuals or that it needs longer to have a sufficient impact to be detectable in species spatial patterns. Instead, there is evidence that the lack of stability in recruit maps could be due to a lack of stability in environmental conditions. An indication for this is that, in recruits, the species in the sub-communities are much more stable than the sub-community maps (and slightly more stable in juveniles). In particular, for small numbers of clusters k one dominating factor in the clustering of recruits is shade-tolerance, and light-gaps in the canopy will open and close between censuses and thereby fundamentally change the spatial pattern of favourable habitats for shade-intolerant species. But other environmental conditions, such as water availability, will also differ between years and therefore change which spatial region in the forest plot is preferable for saplings of a certain species. The juvenile and adult populations, in contrast to the recruits, are made up of multiple recruit cohorts and have averaged out some of the year-to-year changes in environmental conditions. Notably, the clustering of the species based on the data from the recruits also does not show the general trend observed in juveniles and adults that results are more stable over shorter time scales. This could again be interpreted as a lack of a strong spatial association between adult and recruit populations. We also want to highlight the fact that the least similar species clustering for the adults (over the maximum 25-year time-period between 1985 and 2010) is about as good as the match between the clusterings for the recruit cohorts. This could indicate that 25 years is a time-period long enough for survivors to become increasingly rare and for the adult cohorts to become as distinct from one-another as the recruit cohorts already are at shorter time scales. Because the cluster stability for recruits is relatively stable we do not expect the cluster stability for the adults to decrease any further. The stability of the sub-community maps for juveniles and adults, however, is still greater than that for the recruits even over the longest time-period. It will be interesting to see, whether the stability of the sub-community maps will continue to get worse for longer time-periods (once further census measurements will become available), or if it will stabilise at a level above the very low value obtained in the recruits. If the

matching quality stabilises this could indicate that, notwithstanding the temporary changes in environmental conditions that lead to the differences between different recruit cohorts, in the long-term the ecosystem and the environment are stable and community patterns are not changing systematically. If, however, the stability of the sub-community maps continues to decline this could be an indication for long-term shifts in the ecosystem (Feeley *et al.*, 2011) or external factors, such as changing climatic conditions.

The results on the stability of sub-communities and sub-community maps discussed above need to be treated with some caution, because even though our measure of stability is one of relative stability compared to a random null-model, i.e., values for the stability larger than zero are more stable than expected under the null hypothesis of no consistency, this does not give us a measure of the significance of that deviation from zero. It is however possible to compute the same stability measure for randomised sub-community maps and randomised assignments of species to sub-communities (as we compute them for the calculation of the mean expected overlap H_R and S_R). Preliminary results using this approach are presented in Appendix 7.5. These confirm that the results presented are indeed significantly different from the null hypothesis of no consistency between censuses, except, as discussed, for the sub-community maps in recruits. Interestingly, although the relative stability of sub-community maps in the adults is higher than in the juveniles, the stability of the random torus-translated sub-community maps based on the adult maps is also higher than the stability of the random maps based on the juvenile data. This suggests that the deviation from the expected stability under the null hypothesis is not in fact larger for the adults. The reason for this is probably that in the juvenile sub-community maps the areas dominated by the same sub-communities are less spatially contiguous, because canopy gaps with a dominance of shade-intolerant pioneers are sprinkled over the whole area. This might lead to less good matches between torus-translated maps. The comparison between the stability measures based on different data sets is an area that will need further investigations in future work.

As the specific spatial sub-community maps are strongly dependent on the local topography (see Chapter 4) a direct comparison between forest plots might not be possible. Yet, if two forest plots have a large overlap in their species composition it would still be possible to compare the clustering of those species and investigate if they form the same sub-communities. In Chapter 4 we introduced a method to prove that there are non-random sub-communities of species by using a torus-translation null-model. However, this does not provide any information about which species are actually spatially co-associated because of species interactions or environmental conditions and which co-associations are merely coincidental. To investigate this question, exploring the stability of the species clustering across different forests could be a promising avenue.

In conclusion, we believe that the methods presented in this chapter allow us to further study the structure of the normalised co-association matrix and in particular to compare mul-

tiple co-association matrices derived from different subsets of species, different time points or life-stages or even different forests or ecosystems. While we compared patterns from juveniles and adults, the focus of the previous chapter was to study an individual co-association matrix and to extract useful information from it. In this chapter we focused on methods and visualisations that facilitate the comparison between different co-association matrices. We believe this may further improve our understanding of the patterns that we observe, and we have named some of the possibilities this affords and shown some exemplary results. The logical next step will be to apply our methods to comparisons across different forest plots in the CTFS plot network (CTFS, 2013), and develop more rigorous statistical tests for the comparison of the results based on different co-association matrices.

Bibliography

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. 2010. Kernel density estimation via diffusion. *Annals of Statistics*, **38**(5), 2916–2957.

Comita, Liza S, Condit, Richard, & Hubbell, Stephen P. 2007. Developmental changes in habitat associations of tropical trees. *Journal of Ecology*, **95**(3), 482–492.

Comita, Liza S., Muller-Landau, Helene C., Aguilar, Salomon, & Hubbell, Stephen P. 2010. Asymmetric Density Dependence Shapes Species Abundances in a Tropical Tree Community. *Science*, **329**(5989), 330–332.

Condit, Richard, Aguilar, Salomon, Hernandez, Andres, Perez, Rolando, Lao, Suzanne, Angehr, George, Hubbell, Stephen P, & Foster, Robin B. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Nino dry season. *Journal of Tropical Ecology*, **20**(1), 51–72.

Condit, Richard, Engelbrecht, Bettina MJ, Pino, Delicia, Pérez, Rolando, & Turner, Benjamin L. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. *Proceedings of the National Academy of Sciences*, **110**(13), 5064–5068.

CTFS. 2013. *Center for Tropical Forest Science*. [online] http://www.ctfs.si.edu. Accessed: 2013-06-15.

Feeley, Kenneth J., Davies, Stuart J., Perez, Rolando, Hubbell, Stephen P., & Foster, Robin B. 2011. Directional changes in the species composition of a tropical forest. *Ecology*, **92**(4), 871–882.

Harms, Kyle E., Condit, Richard, Hubbell, Stephen P., & Foster, Robin B. 2001. Habitat Associations of Trees and Shrubs in a 50-Ha Neotropical Forest Plot. *Journal of Ecology*, **89**(6), 947–959.

Hubbell, Stephen P., Condit, Richard, & Foster, Robin B. 2005. *Barro Colorado Forest Census Plot Data*. [online] https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Kanagaraj, Rajapandian, Wiegand, Thorsten, Comita, Liza S., & Huth, Andreas. 2011. Tropical tree species assemblages in topographical habitats change in time and with life stage. *Journal of Ecology*, **99**(6), 1441–1452.

Kress, W. John, Erickson, David L., Jones, F. Andrew, Swenson, Nathan G., Perez, Rolando, Sanjur, Oris, & Bermingham, Eldredge. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences*, **106**(44), 18621–18626.

Leigh Jr, Egbert G, Windsor, DM, Rand, A Stanley, & Foster, Robin B. 1990. The impact of the El Niño drought of 1982-83 on a Panamanian semideciduous forest. *Elsevier oceanography series*, **52**, 473–486.

Mayfield, Margaret M, & Levine, Jonathan M. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, **13**(9), 1085–1093.

Muller-Landau, Helene C., & Hardesty, Britta D. 2005. Seed dispersal of woody plants in tropical forests: concepts, examples and future directions. *Pages 267–309 of:* Burslem, Pinard, & Hartley (eds), *Biotic interactions in the Tropics*. Cambridge: Cambridge University Press.

Pearse, William David, Jones, Andy, & Purvis, Andy. 2013. Barro Colorado Island's phylogenetic assemblage structure across fine spatial scales and among clades of different ages. *Ecology*, **[in press]**.

Schnitzer, Stefan A, & Carson, Walter P. 2001. Treefall gaps and the maintenance of species diversity in a tropical forest. *Ecology*, **82**(4), 913–919.

Swenson, Nathan G, Enquist, Brian J, Thompson, Jill, & Zimmerman, Jess K. 2007. The influence of spatial and size scale on phylogenetic relatedness in tropical forest communities. *Ecology*, **88**(7), 1770–1780.

Wiegand, Thorsten, Huth, Andreas, Getzin, Stephan, Wang, Xugao, Hao, Zhanqing, Gunatilleke, CV Savitri, & Gunatilleke, IAU Nimal. 2012. Testing the independent species arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1741), 3312–3320.

Wright, S Joseph, Muller-Landau, Helene C, Condit, Richard, & Hubbell, Stephen P. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology*, **84**(12), 3174–3185.

Chapter 6

Discussion

6.1 Looking back

We started working on the projects presented in this thesis by trying to answer the question: why are rare tree species more aggregated? Species abundance distributions normally show few abundant species and many more rare species in most ecosystems (McGill *et al.*, 2007), and rare species, without any doubt, are an important part of diverse tropical forests. Now we might assume that there are reasons why particular species are rare, and some of the potential reasons for their low abundance might at the same time explain their high aggregation. For example species might be rare because they have such specific habitat requirements that they are only able to survive in a small part of the environment. Alternatively they might be generally weak competitors because of their poor means of seed dispersal. Both those reasons could explain the low abundance and the strong aggregation of rare species. However, using an IBM we could show that even without assuming any such systematic differences between rare and common species, simply through the process of local seed dispersal, we would expect to find this relationship between abundance and aggregation. Even though our IBM was never intended to be used for precise quantitative predictions, we could show that a very simple birth-death process with a range of realistic parameter values for mean seed dispersal distances could fully explain the range of aggregation values for rare and common species observed at BCI (see Figure 3.2). This result is comparable to, and consistent with, Hubbell's neutral theory (Hubbell, 2001), which models the process of speciation and species extinction under the assumption that species are identical, and extinction and survival are pure chance events. Hubbell could show that species differences and niche differentiations are often not needed to explain the observed level of biodiversity in ecosystems. Although there is an ongoing scientific debate about the importance of niche versus neutral processes and the merit of the neutral theory in explaining the biodiversity found in nature (Rosindell *et al.*, 2012; Clark, 2012), neutral theory does obviously not prove that there are no differences between species. Of course, tree species vary widely in their life-history strategies (Loehle, 2000), means of seed dispersal (Muller-Landau & Hardesty, 2005), and tolerance to soil chemicals and drought resistance (Condit *et al.*, 2013), to name just a few of the relevant factors. Sim-

ilarly, having shown using the IBM, that the observed differences in within-species aggregation could solely be explained by the abundance of species and the fact that most saplings are close to their parent, does not rule out that other factors may also affect the spatial patterns of trees. Instead, having established the link between abundance and aggregation we can now account for this relationship when studying other factors that potentially affect spatial aggregation.

In Chapter 2 we built on our results on abundance and aggregation to further examine the relationship between breeding system and aggregation. We showed that contrary to the relationship between abundance and aggregation that was both predicted from our model and confirmed in the data from BCI, the predicted effect of dioecy is not found in the forest at BCI. Such a negative result, however, can be just as informative as a positive result: Not having found the relationship between abundance and dioecy means that most dioecious species at BCI must have developed mechanisms to decrease their aggregation rather than developed means to overcome the disadvantage that a more aggregated spatial distribution can bring. The mechanisms by which dioecious species manage to overcome their competitive disadvantages might be different between different dioecious species (Queenborough *et al.*, 2007) and more information on the sex of individuals within the CTFS forest plots will be needed to use the available data on species spatial pattern to answer more specific questions on breeding systems in trees.

In Chapter 3 we looked at the relationship of within-species aggregation and changes of abundance over time. When we analyse spatial patterns to gain information on the processes that have shaped those patterns, we necessarily always look back into the past of the system we are studying. It therefore suggests itself to use spatial analyses not just to study processes governing the ecosystem in general, but also to explicitly try to gain information on the state of the system in the past and how it has changed. We could argue that changes are often what is of most practical interest when studying the environment, because changes might be causes for concern for the stability of the system, potential threats (or opportunities), and therefore reasons for us to intervene or prepare for the change. Changes in the recent past are often the best predictors for further changes in the future. Being able to infer past changes from current pattern will therefore be an important skill for the developing field of predictive ecology that aims to forecast changes in ecosystems under changing conditions (Evans *et al.*, 2012).

In the second half of this thesis we have studied multivariate pattern of spatial co-associations. Multivariate spatial patterns are influenced by many different factors, just as the spatial distribution of individual species. Consequently, just as we had to correct for abundance when studying other factors that might be related to within-species aggregation, we also have to correct for factors influencing the bi-variate species co-associations that are not the ones we are examining. In Chapter 4 we introduced a new method to normalise co-association values between pairs of species that take into account the abundance and within-

species spatial patterns of the species. This method makes the normalised co-association values comparable between different pairs of species and therefore allows us to ask questions about the joint co-association values between all pairs of species in an ecosystem. Because of the many rare species, and because species patterns are strongly auto-correlated, the information in the spatial distribution of a single tree species from the tropical rainforest is often very limited. However, if we can manage to extract information from the whole matrix of all cross-species interactions, this provides a potentially much richer data set. We have demonstrated how this information could be used to create a map of the structure of the ecosystem, which provided us with information on habitats and the importance of disturbances and canopy gaps (Chapter 4). We then outlined some methods to compare the structure of the normalised co-associations matrix between different data-sets and analysed its stability over time or through life-stages of individuals (Chapter 5). Multivariate spatial pattern analyses have only in the recent years become more widespread in ecology, and we would argue that we have only begun to scratch the surface of a technique whose potential is still to be uncovered.

6.2 Going forward

In this final section we want to address three different aspects of future work. First, questions and proposals for future research that arise directly from the work presented in this thesis. Second, future developments that we expect will shape the direction in which tropical forest ecology will progress over the coming years. And finally, the opportunities for theoretical ecology to engage citizen scientists.

Follow-ups

The main focus of the work presented in this thesis was to prove the usefulness and applicability of the methods presented here. Using the IBM we made predictions of the effects we would expect to find in the BCI data, but the process modelled in the IBM was deliberately simple. The basic mechanisms of birth, dispersal, and death of individuals can be found in many ecosystems, and testing our hypotheses on the relationship between abundance, recent change in abundance, and aggregation in different ecosystems would therefore be the next logical step. Our methods can be directly applied to any of the CTFS forest plots, but also to entirely different ecosystems such as coral reefs (Karlson *et al.*, 2007). The same is true for our work with the normalised co-association matrix. We used the data from BCI to show how it could be used to find and analyse structure in an ecosystem, but the main advantage of the normalisation is that different co-association matrices based on different data sets become comparable. However, there are open questions that need to be answered. While it is relatively simple to develop random null-models to decide if a pattern significantly deviates from a null-model, it is more difficult to compare two different unique patterns found in nature and decide if the two patterns are systematically different from one-another or just show random variations. In principle there are at least two different approaches we might want to take to

tackle this problem. We could develop better models, use the empirical data we have to fit the parameters of the model and then run multiple instances of the model to produce a range of outcomes, which would give us an estimation of the variability in the modelled process. For this approach to succeed we need sufficient confidence that our model is actually a close enough match of the processes that created the empirical patterns we want to compare. Alternatively, we could repeat our analysis multiple times on only parts of the empirical data, for example if we want to compare two 50 ha forest plots we could try to split each plot into fifty 1 ha plots and then use statistical tests to determine if the results of the 1 ha sub-plots from one site differ significantly from the results obtained from the other site. This approach is obviously very data hungry as effectively the available data has to be divided by the desired number of repetitions.

A further research path that leads on from the work discussed here would be to apply the methods developed in the second half of this thesis to ask some of the questions dealt with in Chapter 3. We could use the normalised co-association matrix to try and extract information from the past of an ecosystem to make predictions about its future. The IBM developed for the analyses in Chapter 2 and 3 could be expanded to incorporate species interactions and sub-communities with different properties. We could then use the IBM to make specific predictions on how distinct events like canopy disturbances or droughts would affect the forest structure and look for evidence for those hypotheses in the normalised co-association matrix.

Coming-up

With climate change high-up on the global agenda, forests are increasingly appreciated as carbon stocks and carbon sinks and their protection has become the topic of international diplomacy. The United Nations REDD initiative (Gibbs *et al.*, 2007) aims to use market mechanisms to pay for the protection of forests. Expanding the CTFS network of tropical forest plots (CTFS, 2013) into the Smithsonian Institution Global Earth Observatories (SIGEO) forest research network has the aim to integrate measurements of carbon fluxes into the CTFS protocol and to expand the network beyond the tropics into temperate forests. This research initiative thereby aims to better quantify the role that forests play in the climate system and how they are affected by climate change. Understanding the dynamics in forests with regard to carbon budgets and the value and stability of degraded forests, such as palm plantations and reforestation projects, are likely to be the main areas of interest in forest community ecology over the coming years. Measures such as our normalised co-association matrix could potentially be used to classify forests into different categories describing the forest status, such as *old-growth*, *under stress*, or *recovering*, which could be used to judge the success of forest management strategies and ascribe the forest value under climate change mitigation schemes.

More forests being mapped in detail under the CTFS/SIGEO initiative or similar efforts, e.g., rainfor (rainfor, 2013) and AfriTRON (AfriTRON, 2013), will provide valuable addi-

tions to the data pool for comparative analyses. As additional census data becomes available from long-running projects such as the forest dynamics research plot at BCI, even better analyses of temporal changes will be made possible, such as for example the effects of singular events like years of strong El Niño droughts and shifts in forest structure caused by longer-term climatic changes. But this increase in available data is mainly an incremental gain. Fundamental shifts in our ability to monitor and analyse forest ecosystems are more likely to come from two technological innovations that are currently starting to have an impact, remote sensing and DNA analysis.

Remote sensing from satellites or air planes is already used to establish the extent of forests and to monitor global forest loss (Gaveau *et al.*, 2007; Hansen *et al.*, 2010). However, recent studies have also used remote sensing to distinguish successional stages and types of vegetation (Castillo *et al.*, 2012; Gairola *et al.*, 2013). Holmgren & Persson (2004) used airborne laser scanning data to identify the species identity of trees in a Norwegian forest with three tree species and achieved over 95% accuracy. Garzon-Lopez *et al.* (2012) explored the use of aerial photographs to detect selected canopy species at BCI, and even though they could detect only 40% of the individuals of their target species, the collected data was sufficient to detect within-species aggregation patterns of species. Further improvements in remote sensing technology and data analysis techniques might allow us to map spatial patterns of at least some species or groups of species over a large, or even global, scale – even in highly diverse tropical rainforests. Though that data might be low quality compared to ground-based surveys in the foreseeable future, the amount of data collected at relatively low costs might more than compensate for the lower accuracy.

While remote sensing might replace ground-based surveys for some research questions in the future, DNA analysis promises to be the ideal addition to mapped forest plots. In Jones *et al.* (2005) the authors used DNA analyses to determine the parent trees for seeds of the species *Jacaranda copaia* they collected at the BCI forest plot. With those methods falling in price and becoming more easily available, this method could be extended to genotype all trees in the forest plot. Knowing the parental-child relationships between all trees in the forest plot will enable us to much better model and understand seed dispersal, seedling establishment and pollination. A map of genetic relationships of individuals would in particular inform our understanding of the differences between dioecious and hermaphrodite species.

Joining-in

In 1999 the SETI@home project was launched, asking people from around the world to donate computing power to search for signals of extraterrestrial life in radio telescope data from space (Korpela *et al.*, 2001). Computers connected to the internet could run the software provided by the researchers and, when idle, analyse the radio-telescope data sent to them via the internet and report back the results. In that way a massive virtual supercomputer was created allowing a deeper search for signals in the data than would otherwise have been possible.

Since then multiple similar distributed computing projects have been started allowing people to share their computer resources for scientific research. A different approach to draw on the resources of large numbers of volunteers, that has gained considerable attention over the last couple of years, has been termed *citizen science* (Henderson, 2012). The eBird project for example has developed a web-platform that allows volunteer birdwatchers to report sights of birds (Sullivan *et al.*, 2009). Nowadays, an increasing number of people carry smartphones that include cameras, GPS sensors and an internet connection. Thus, large numbers of people can report animal or plant sightings together with photos and precise geo-location information, making citizen science a valuable resource to monitor species distributions. Besides the information transmitted from the citizen scientist to the research establishment, citizen involvement also often explicitly aims to facilitate an information flow in the opposite direction, educating the public about the scientific process as well as the subject they become involved in helping with.

In ecology all attempts of citizen involvement we know of are about gathering field data. To our knowledge there are currently no theoretical ecology citizen science or distributed computing projects. Most computer models in ecology do not require large enough resources to benefit from the computing power provided by distributed computing initiatives. Using volunteers to collect data was the most obvious step for citizen engagement in ecology. However, we believe that, if for no other reason than for the purpose of science communication, involving interested members of the public in the research in theoretical ecology would be a worthwhile endeavour. One way of bringing citizen science into the field of theoretical ecology could be through *serious games*. In serious games the medium of computer games is used to provide the user with a learning experience or to let them solve real problems using the motivation provided by a gaming framework. Translated to theoretical forest ecology this could for example mean that users could create an artificial tree species, determine the traits of their species under constraints set by the game rules that would enforce biologically realistic trade-offs and prohibit to create a super-species, and then let their species compete against species of other players in an individual-based model. Players could be allowed to let their species compete in different environments or maybe even to intervene in the simulation, attacking their opponent's species with droughts and wildfires. The data created by such a game could be analysed in terms of the numbers of species that could co-exist under different conditions or the traits that are favoured under certain circumstances. In such a model theoretical ecologists could make use of both the computing resources of the citizen scientist running the simulation model, and the problem solving skills of players to find ways for species to co-exist and survive in difficult conditions. And all this would be possible while users are being educated about concepts of species co-existence, trait trade-offs, and the use of computer models in ecological research.

Bibliography

AfriTRON. 2013. *African Tropical Rainforest Observation Network*. [online] http://www.geog.leeds.ac.uk/projects/afritron/. Accessed: 2013-06-15.

Castillo, Mauricio, Rivard, Benoit, Sánchez-Azofeifa, Arturo, Calvo-Alvarado, Julio, & Dubayah, Ralph. 2012. LIDAR remote sensing for secondary Tropical Dry Forest identification. *Remote Sensing of Environment*, **121**, 132–143.

Clark, James S. 2012. The coherence problem with the Unified Neutral Theory of Biodiversity. *Trends in ecology & evolution*, **27**(4), 198–202.

Condit, Richard, Engelbrecht, Bettina MJ, Pino, Delicia, Pérez, Rolando, & Turner, Benjamin L. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. *Proceedings of the National Academy of Sciences*, **110**(13), 5064–5068.

CTFS. 2013. *Center for Tropical Forest Science*. [online] http://www.ctfs.si.edu. Accessed: 2013-06-15.

Evans, Matthew R, Norris, Ken J, & Benton, Tim G. 2012. Predictive ecology: systems approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1586), 163–169.

Gairola, Sanjay, Procheş, Şerban, & Rocchini, Duccio. 2013. High-resolution satellite remote sensing: a new frontier for biodiversity exploration in Indian Himalayan forests. *International Journal of Remote Sensing*, **34**(6), 2006–2022.

Garzon-Lopez, Carol X, Bohlman, Stephanie A, Olff, Han, & Jansen, Patrick A. 2012. Mapping Tropical Forest Trees Using High-Resolution Aerial Digital Photographs. *Biotropica*.

Gaveau, David LA, Wandono, Hagnyo, & Setiabudi, Firman. 2007. Three decades of deforestation in southwest Sumatra: Have protected areas halted forest loss and logging, and promoted re-growth? *Biological Conservation*, **134**(4), 495–504.

Gibbs, Holly K, Brown, Sandra, Niles, John O, & Foley, Jonathan A. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, **2**(4), 045023.

Hansen, Matthew C, Stehman, Stephen V, & Potapov, Peter V. 2010. Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences*, **107**(19), 8650–8655.

Henderson, Sandra. 2012. Citizen science comes of age. *Frontiers in Ecology and the Environment*, **10**(6), 283–283.

Holmgren, Johan, & Persson, Åsa. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, **90**(4), 415–423.

Hubbell, Stephen P. 2001. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. Princeton University Press.

Jones, FA, Chen, J, Weng, G-J, & Hubbell, SP. 2005. A genetic evaluation of seed dispersal in the neotropical tree Jacaranda copaia (Bignoniaceae). *The American Naturalist*, **166**(5), 543–555.

Karlson, Ronald H., Cornell, Howard V., & Hughes, Terence P. 2007. Aggregation influences coral species richness at multiple spatial scales. *Ecology*, **88**(1), 170–177.

Korpela, Eric, Werthimer, Dan, Anderson, David, Cobb, Jeff, & Leboisky, Matt. 2001. Seti@home-massively distributed computing for seti. *Computing in science & engineering*, **3**(1), 78–83.

Loehle, Craig. 2000. Strategy space and the disturbance spectrum: A life-history model for tree species coexistence. *The American Naturalist*, **156**(1), 14–33.

McGill, Brian J, Etienne, Rampal S, Gray, John S, Alonso, David, Anderson, Marti J, Benecha, Habtamu Kassa, Dornelas, Maria, Enquist, Brian J, Green, Jessica L, He, Fangliang, *et al.* 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology letters*, **10**(10), 995–1015.

Muller-Landau, Helene C., & Hardesty, Britta D. 2005. Seed dispersal of woody plants in tropical forests: concepts, examples and future directions. *Pages 267–309 of:* Burslem, Pinard, & Hartley (eds), *Biotic interactions in the Tropics*. Cambridge: Cambridge University Press.

Queenborough, Simon A., Burslem, David F. R. P., Garwood, Nancy C., & Valencia, Renato. 2007. Determinants of biased sex ratios and inter-sex costs of reproduction in dioecious tropical forest trees. *Am. J. Bot.*, **94**(1), 67–78.

rainfor. 2013. *Amazon Forest Inventory Network*. [online] http://www.rainfor.org/. Accessed: 2013-06-15.

Rosindell, James, Hubbell, Stephen P, He, Fangliang, Harmon, Luke J, & Etienne, Rampal S. 2012. The case for ecological neutral theory. *Trends in ecology & evolution*, **27**(4), 203–208.

Sullivan, Brian L, Wood, Christopher L, Iliff, Marshall J, Bonney, Rick E, Fink, Daniel, & Kelling, Steve. 2009. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**(10), 2282–2292.

Chapter 7

Appendices

7.1 Species list for Chapter 3

Name	$\log_{10}(1+\Omega_{0,10})$	N_{2005}	$\Delta_{20}\log(N)^a$	$\Delta_5\log(N)^b$	gf^c
Acalypha diversifolia	0.979	746	-0.20717	0.18254	S
Adelia triloba	1.2249	142	-0.34602	-0.059899	U
Alchornea costaricensis	0.98997	227	-0.1409	-0.001909	T
Alibertia edulis	0.59435	370	0.034176	0.016752	U
Allophylus psilospermus	0.6628	103	-0.22016	-0.036381	M
Alseis blackiana	0.43402	7754	-0.016324	-0.0065593	T
Anaxagorea panamensis	2.4331	794	0.22496	0.025919	S
Andira inermis	0.3494	278	-0.043094	-0.0092735	T
Annona acuminata	0.99463	485	-0.03276	-0.011488	S
Apeiba membranacea	0.72168	289	-0.071857	0.031146	T
Aspidosperma spruceanum	0.60315	469	-0.00092502	-0.0091627	T
Astrocaryum standleyanum	0.40139	165	-0.148	-0.056674	M
Beilschmiedia pendula	0.62382	2116	-0.10051	-0.039785	T
Brosimum alicastrum	0.42098	892	-0.00097266	-0.0029115	T
Calophyllum longifolium	0.87149	1427	0.29709	0.10096	T
Capparis frondosa	0.4342	2749	-0.12573	-0.036351	S
Casearia aculeata	0.44317	434	-0.041938	0.01424	U
Casearia arborea	0.67452	124	-0.25683	-0.049593	T
Casearia sylvestris	0.36142	131	-0.24822	-0.061706	M
Cassipourea elliptica	0.64438	1069	0.10058	0.025948	M
Cecropia insignis	1.1801	1144	0.41399	0.20351	T
Cecropia obtusifolia	1.5092	235	0.79128	0.36247	M

Continued on Next Page...

Table 7.1 – Continued

Name	$\log_{10}(1+\Omega_{0,10})$	N_{2005}	$\Delta_{20}\log(N)^a$	$\Delta_5\log(N)^b$	gf^c
Celtis schippii	0.34726	114	-0.1632	-0.029455	M
Cestrum megalophyllum	0.50286	54	-0.64235	0.033424	S
Chamguava schippii	1.3809	449	0.27385	0.071321	U
Chrysochlamys eclipes	0.70771	391	-0.04731	-0.0087961	S
Chrysophyllum argenteum	0.36925	670	0.14393	-0.0025851	T
Chrysophyllum cainito	0.70965	147	0.26969	0.040213	T
Coccoloba coronata	0.17826	112	-0.23182	-0.01148	M
Coccoloba manzinellensis	0.54496	372	-0.078792	-0.033683	U
Conostegia cinnamomea	1.0258	98	-0.45593	-0.10568	S
Cordia alliodora	1.2792	148	0.13684	0.19254	T
Cordia bicolor	0.59558	658	-0.062016	-0.065436	M
Cordia lasiocalyx	0.41719	1171	-0.1539	-0.040008	M
Coussarea curvigemmia	0.68563	2058	0.091249	-0.0039911	U
Croton billbergianus	1.5694	468	-0.12285	0.11636	U
Cupania seemannii	0.52074	1213	0.067245	0.00035818	U
Dendropanax arboreus	0.63229	88	-0.25964	-0.068355	T
Desmopsis panamensis	0.34898	11327	-0.029459	0.0014979	U
Drypetes standleyi	0.67677	2180	-0.016995	-0.0025821	T
Erythrina costaricensis	0.5329	89	-0.42168	-0.10369	U
Erythroxylum macrophyllum	0.41098	242	-0.10048	-0.0053507	M
Erythroxylum panamense	0.36461	110	0.016087	0.032793	U
Eugenia coloradoensis	0.50858	611	-0.13927	-0.028872	T
Eugenia galalonensis	0.5465	1751	0.17845	0.043257	U
Eugenia nesiotica	0.5681	482	-0.050952	-0.02707	M
Eugenia oerstediana	0.55935	1816	-0.08449	-0.025991	M
Faramea occidentalis	0.37024	26038	0.015311	-0.011716	U
Garcinia intermedia	0.41166	4602	0.059369	0.019001	M
Garcinia madruno	0.55591	393	-0.24256	0.021521	M
Guapira standleyana	0.5117	164	-0.10738	-0.0026401	T
Guarea ‘fuzzy’	0.37257	823	-0.25808	-0.10038	M/T
Guarea guidonia	0.46459	1774	-0.013023	-0.014446	M
Guatteria dumetorum	0.45206	896	-0.23125	-0.064725	T
Guettarda foliacea	0.44291	268	-0.15506	-0.050432	U

Continued on Next Page...

Table 7.1 – Continued

Name	$\log_{10}(1+\Omega_{0,10})$	N_{2005}	$\Delta_{20}\log(N)^a$	$\Delta_5\log(N)^b$	gf^c
Gustavia superba	1.1118	734	-0.047588	-0.010522	M
Hasseltia floribunda	0.61258	484	-0.32034	-0.049154	M
Heisteria acuminata	0.68302	100	-0.049218	-0.025306	U
Heisteria concinna	0.4308	927	-0.014739	-0.012013	M
Herrania purpurea	0.40619	521	-0.0066179	0.0084168	U
Hirtella triandra	0.42425	4566	-0.0086636	-0.01798	M
Hura crepitans	0.39797	103	-0.06271	-0.0083521	T
Hybanthus prunifolius	0.38371	29846	-0.13881	-0.029286	S
Inga acuminata	0.54583	424	0.22253	0.061518	U
Inga goldmanii	0.46007	313	-0.17097	-0.041014	T
Inga marginata	0.82541	400	-0.29557	0.0021769	T
Inga nobilis	0.43406	615	-0.10599	-0.033293	M
Inga pezizifera	0.92523	126	-0.1941	0.021189	T
Inga sapindoides	0.34476	205	-0.24213	-0.084445	M
Inga thibaudiana	1.0393	178	0.36393	0.13983	T
Inga umbellifera	0.41052	797	-0.10028	-0.01762	M
Jacaranda copaia	0.73768	280	-0.057992	0.025554	T
Lacistema aggregatum	0.42229	1276	-0.11635	-0.041826	U
Laetia thamnia	0.75845	410	-0.099866	-0.040429	U
Licania hypoleuca	1.2912	127	0.021018	0	M
Licania platypus	0.61795	266	-0.043466	-0.025367	T
Lonchocarpus heptaphyllus	0.60893	659	-0.10951	-0.047992	T
Luehea seemannii	1.0604	190	0	-0.0557	T
Macrocnemum roseum	1.3747	94	-0.052178	0.00934	M
Maquira guianensis	0.42212	1396	-0.014381	-0.019467	M
Miconia affinis	0.98904	389	-0.0022272	0.015918	U
Miconia argentea	0.84305	518	-0.11626	-0.063821	M
Miconia nervosa	0.97495	262	-0.048566	-0.048566	S
Mosannona garwoodii	0.43598	472	0.17015	0.035453	M
Mouriri myrtilloides	0.3954	6540	-0.06984	0.0023304	S
Nectandra cissiflora	0.79694	179	-0.2799	-0.041613	T
Nectandra lineata	0.86592	112	0.003895	0.049218	M
Ocotea cernua	0.3933	237	-0.16307	0.026443	M

Continued on Next Page...

Table 7.1 – Continued

Name	$\log_{10}(1+\Omega_{0,10})$	N_{2005}	$\Delta_{20}\log(N)^a$	$\Delta_5\log(N)^b$	gf^c
Ocotea oblonga	0.74888	162	-0.064643	0.039253	T
Ocotea puberula	0.63736	131	-0.27189	-0.016268	T
Ocotea whitei	1.2046	374	-0.39374	-0.058572	T
Oenocarpus mapora	0.42352	1787	0.02142	-0.010327	M
Ouratea lucens	0.57484	1227	-0.0038761	0.012933	S
Palicourea guianensis	1.0229	851	0.10973	-0.0070865	S
Pentagonia macrophylla	0.53661	300	-0.233	-0.011429	U
Perebea xanthochyma	1.0649	233	-0.059155	-0.014661	M
Picramnia latifolia	0.54764	1059	-0.045881	0.0045347	U
Piper cordulatum	1.0487	50	-1.8701	-0.26951	S
Piper reticulatum	0.67299	131	-0.11318	0.027366	U
Platymiscium pinnatum	0.7729	146	-0.068643	-0.0059089	T
Platypodium elegans	0.73966	122	-0.12047	-0.0035453	T
Poulsenia armata	0.68824	1162	-0.36277	-0.081852	T
Pouteria reticulata	0.34523	1204	-0.15465	-0.065502	T
Prioria copaifera	0.81119	1348	-0.017986	-0.010189	T
Protium costaricense	0.46258	698	-0.093161	-0.036958	M
Protium panamense	0.4988	2853	-0.0022774	-0.0070962	M
Protium tenuifolium	0.57166	2829	-0.0019911	-0.0033643	M
Psychotria horizontalis	0.67805	3119	-0.3144	-0.10115	S
Psychotria marginata	0.8461	581	-0.072148	0.021451	S
Pterocarpus rohrii	0.39344	1380	-0.071241	-0.025365	T
Quararibea asterolepis	0.4819	2137	-0.046772	-0.013013	T
Quassia amara	0.91344	119	-0.11197	-0.045027	U
Randia armata	0.36705	958	-0.081216	-0.019935	U
Rinorea sylvatica	1.045	2277	-0.061104	-0.007188	S
Senna dariensis	1.1098	135	-0.0032051	0.31079	S
Simarouba amara	0.53982	1477	0.073166	0.079475	T
Siparuna pauciflora	0.60221	367	0.015663	0.064979	U
Sloanea terniflora	0.48244	461	-0.11082	-0.029146	T
Socratea exorrhiza	0.94901	540	-0.13507	-0.061397	M
Sorocea affinis	0.38903	2539	-0.12297	-0.04789	S
Spondias radlkoferi	1.2127	265	0.1827	0.0082725	T

Continued on Next Page...

Table 7.1 – Continued

Name	$\log_{10}(1+\Omega_{0,10})$	N_{2005}	$\Delta_{20}\log(N)^a$	$\Delta_5\log(N)^b$	gf^c
Stylogyne turbacensis	0.40823	691	-0.028578	0.0076083	*
Swartzia simplex	0.34937	5710	0.036311	0.01203	U
Symphonia globulifera	0.94709	152	-0.071009	-0.022276	T
Tabebuia rosea	0.61584	230	-0.10956	-0.027438	T
Tabernaemontana arborea	0.4246	1593	0.076084	0.024103	T
Tachigali versicolor	0.51173	2234	-0.12484	-0.046243	T
Talisia nervosa	0.93121	722	-0.053153	-0.014783	U
Talisia princeps	0.45786	664	0.017347	0.021451	M
Tetragastris panamensis	0.46957	4493	0.081061	0.026196	T
Trichilia pallida	0.5128	478	-0.077968	-0.033222	M
Trichilia tuberculata	0.47944	11344	-0.06482	-0.023799	T
Triplaris cumingiana	0.67672	242	-0.15526	-0.027804	M
Trophis caucana	0.90447	149	-0.39969	-0.069852	U
Trophis racemosa	0.51687	253	-0.10876	-0.051724	M
Unonopsis pittieri	0.53173	621	-0.10233	-0.024474	M
Virola sebifera	0.41939	1394	-0.21234	-0.063372	M
Virola surinamensis	0.60655	183	-0.16079	-0.040745	T
Xylopia macrantha	0.79481	1414	0.18761	0.064446	M
Xylosma oligandra	0.97814	67	-0.39664	-0.15165	S
Zanthoxylum acuminatum	0.92224	93	-0.35785	-0.095975	T
Zanthoxylum ekmanii	1.1028	194	-0.1084	-0.081414	T
Zanthoxylum panamense	1.1866	178	-0.1334	0.095084	T

Table 7.1: List of species used in Chapter 3

[a]: $\Delta_{20}\log(N)$ denotes the difference of $\log_{10}(N_t)$ and $\log_{10}(N_{t-\Delta t})$ in the 20-year interval between 1985 and 2005.

[b]: $\Delta_5\log(N)$ denotes the difference of $\log_{10}(N_t)$ and $\log_{10}(N_{t-\Delta t})$ in the 5-year interval between 2000 and 2005.

[c]: Growth form; S = shrub; U = understory tree; M = mid-canopy tree; T = top-canopy tree; * = missing data.

7.2 Species list for Chapter 4

Name	N_{2010}	# of adults	# of juveniles	ST[a]	SR[b]
Acacia melanoceras	48	17	31	N/A	40
Acalypha diversifolia	1023	582	441	N/A	20
Adelia triloba	143	65	78	-0.67	100
Aegiphila panamensis	40	29	11	-0.47	40
Alchornea costaricensis	316	102	214	-2.58	200
Alibertia edulis	417	68	349	0.62	40
Allophylus psilospermus	111	62	49	-0.52	40
Alseis blackiana	7928	577	7351	1.11	200
Annona spraguei	149	14	135	-3.4	80
Apeiba membranacea	308	97	211	-1	300
Apeiba tibourbou	50	24	26	-2.34	80
Aspidosperma spruceanum	473	31	442	1.28	300
Astronium graveolens	117	21	96	0.53	300
Beilschmiedia pendula	1996	114	1882	0.63	300
Brosimum alicastrum	872	49	823	1.15	300
Calophyllum longifolium	1823	13	1810	-0.03	300
Casearia aculeata	452	78	374	1.12	50
Casearia arborea	131	41	90	-0.32	200
Casearia sylvestris	128	38	90	-0.07	100
Cassipourea elliptica	1108	154	954	1.09	80
Cavanillesia platanifolia	36	16	20	-0.24	500
Cecropia insignis	891	82	809	-4.77	300
Cecropia obtusifolia	264	150	114	-4.27	80
Ceiba pentandra	62	28	34	-1.67	600
Celtis schippii	108	20	88	0.81	160
Chamguava schippii	541	92	449	0.61	40
Chrysochlamys eclipes	406	282	124	N/A	20
Chrysophyllum argenteum	711	15	696	1.05	300
Chrysophyllum cainito	171	15	156	0.36	300
Coccoloba coronata	138	23	115	1.42	80
Coccoloba manzinellensis	351	11	340	2.07	100
Cordia alliodora	189	38	151	-2.14	200

Continued on Next Page...

Table 7.2 – Continued

Name	N_{2010}	# of adults	# of juveniles	ST[a]	SR[b]
Cordia bicolor	681	268	413	-0.79	160
Cordia lasiocalyx	1143	214	929	-0.26	100
Coussarea curvigemmia	2010	892	1118	0.69	30
Croton billbergianus	631	158	473	-3.79	50
Cupania seemannii	1271	209	1062	1.69	50
Dendropanax arboreus	79	19	60	0	300
Desmopsis panamensis	11040	3383	7657	0.65	30
Dipteryx oleifera	52	30	22	0.73	300
Drypetes standleyi	2110	107	2003	1.17	200
Erythrina costaricensis	68	43	25	N/A	50
Erythroxylum macrophyllum	284	39	245	0.06	40
Eugenia galalonensis	1975	139	1836	1.16	40
Eugenia nesiotica	502	48	454	1.52	100
Eugenia oerstediana	1838	38	1800	0.15	200
Faramea occidentalis	24989	9459	15530	0.66	50
Garcinia intermedia	4817	143	4674	1.13	100
Guapira standleyana	155	50	105	0.82	300
Guarea bullata	725	70	655	0.21	80
Guarea guidonia	1889	734	1155	0.56	40
Guatteria dumetorum	876	53	823	-0.13	300
Guazuma ulmifolia	74	28	46	N/A	300
Guettarda foliacea	252	59	193	0.65	100
Gustavia superba	692	607	85	0.99	100
Hampea appendiculata	191	36	155	-2.72	80
Hasseltia floribunda	418	201	217	0	80
Heisteria acuminata	94	46	48	0.73	50
Heisteria concinna	895	196	699	0.7	150
Hieronyma alchorneoides	118	26	92	-0.4	300
Hirtella triandra	4407	1026	3381	0.82	80
Hura crepitans	95	72	23	0.72	300
Inga acuminata	606	88	518	-0.23	80
Inga marginata	767	43	724	-1.7	300
Inga nobilis	557	144	413	0.83	80

Continued on Next Page...

Table 7.2 – Continued

Name	N_{2010}	# of adults	# of juveniles	ST[a]	SR[b]
Inga sapindoides	197	30	167	0.2	160
Inga umbellifera	765	50	715	0.69	80
Jacaranda copaia	327	167	160	-1.94	300
Lacistema aggregatum	1264	135	1129	0.45	50
Lacmellea panamensis	102	39	63	0.59	160
Laetia thamnia	409	51	358	0.04	80
Licania hypoleuca	141	15	126	1.08	80
Lindackeria laurina	53	41	12	-0.34	100
Lonchocarpus heptaphyllus	665	19	646	0.58	300
Luehea seemannii	215	44	171	-1.14	300
Macrocnemum roseum	87	29	58	1.35	80
Maquira guianensis	1315	164	1151	0.88	100
Maytenus schippii	76	31	45	1.13	80
Miconia affinis	437	140	297	-1.29	40
Miconia argentea	675	58	617	-3.3	100
Mosannona garwoodii	502	146	356	1.5	40
Mouriri myrtilloides	6804	2789	4015	N/A	20
Ocotea cernua	288	87	201	0.57	40
Ocotea oblonga	240	21	219	-1.61	300
Ocotea whitei	390	70	320	-0.78	300
Oenocarpus mapora	1802	1418	384	N/A	80
Ouratea lucens	1206	198	1008	N/A	30
Pentagonia macrophylla	306	208	98	0.87	20
Perebea xanthochyma	226	50	176	0.72	80
Picramnia latifolia	1080	222	858	0.42	40
Piper arboreum	24	12	12	0.05	30
Piper reticulatum	136	74	62	0.23	40
Platymiscium pinnatum	164	26	138	-0.6	300
Platypodium elegans	109	17	92	-0.36	300
Posoqueria latifolia	73	20	53	1.4	80
Poulsenia armata	993	61	932	-0.66	300
Pourouma bicolor	130	10	120	-1.58	200
Pouteria reticulata	1084	88	996	0.56	300

Continued on Next Page...

Table 7.2 – Continued

Name	N_{2010}	# of adults	# of juveniles	ST[a]	SR[b]
Pouteria stipitata	60	23	37	0.66	160
Prioria copaifera	1327	90	1237	1.33	600
Protium costaricense	697	27	670	0.13	160
Protium panamense	3020	126	2894	0.95	80
Protium tenuifolium	2900	191	2709	0.63	200
Psychotria grandis	57	16	41	-0.95	30
Quararibea asterolepis	2170	338	1832	1.05	300
Quassia amara	115	71	44	1.12	40
Randia armata	937	472	465	0.84	50
Rinorea sylvatica	2264	1157	1107	N/A	20
Simarouba amara	1577	63	1514	-1.25	300
Siparuna pauciflora	359	83	276	0.29	40
Socratea exorrhiza	500	348	152	N/A	80
Solanum hayesii	68	30	38	-3.18	40
Sorocea affinis	2295	1330	965	N/A	30
Spondias mombin	120	19	101	-1.73	300
Spondias radlkoferi	271	25	246	-1.13	300
Sterculia apetala	53	13	40	0.05	400
Stylogyne turbacensis	706	110	596	N/A	30
Tabebuia guayacan	63	16	47	1.04	300
Tabebuia rosea	234	21	213	-0.19	300
Tabernaemontana arborea	1732	152	1580	0.8	300
Talisia nervosa	703	419	284	2.01	30
Terminalia amazonia	47	16	31	0.58	300
Terminalia oblonga	90	29	61	N/A	300
Tetragastris panamensis	4622	125	4497	0.79	300
Thevetia ahouai	84	64	20	N/A	20
Trattinnickia aspera	79	15	64	N/A	300
Trichilia pallida	472	132	340	0.07	80
Trichilia tuberculata	10841	330	10511	0.61	300
Triplaris cumingiana	230	48	182	-0.65	200
Trophis caucana	136	105	31	N/A	40
Trophis racemosa	228	28	200	0.77	80

Continued on Next Page...

Table 7.2 – Continued

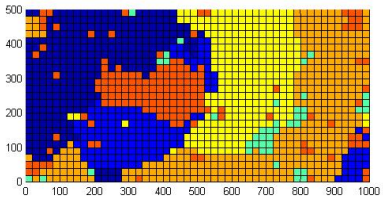
Name	N_{2010}	# of adults	# of juveniles	ST[a]	SR[b]
Unonopsis pittieri	643	217	426	0.13	80
Virola multiflora	41	10	31	0.01	300
Virola sebifera	1274	234	1040	0.47	200
Virola surinamensis	178	87	91	0.04	300
Vismia baccifera	56	39	17	-1.65	20
Xylopia macrantha	1658	221	1437	0.69	100
Xylosma oligandra	53	42	11	N/A	30
Zanthoxylum acuminatum	89	13	76	-0.62	200
Zanthoxylum ekmanii	257	92	165	-3.52	300
Zanthoxylum panamense	241	23	218	-0.58	300

Table 7.2: List of species used in Chapter 4

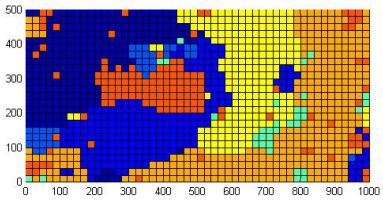
[a]: ST: Shade-tolerance index as defined in Comita et al. (2010).

[b]: SR: Size at reproduction [in mm DBH]. The estimates of size at reproduction are estimates by Robin Foster on the size at which species first reproduce, as provided by the Centre for Tropical Forest Science (CTFS).

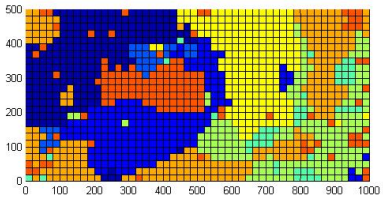
7.3 Sub-community maps for $k = 6$ to $k = 10$



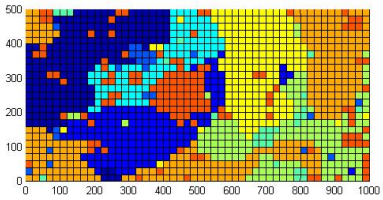
(a) k = 6



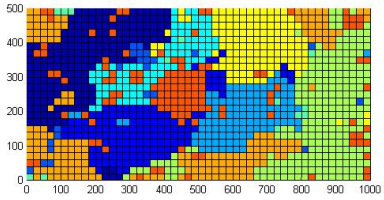
(b) k = 7



(c) k = 8



(d) k = 9

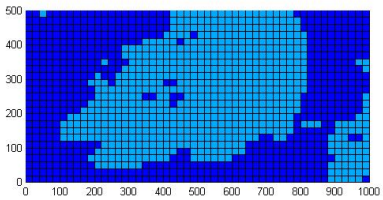


(e) k = 10

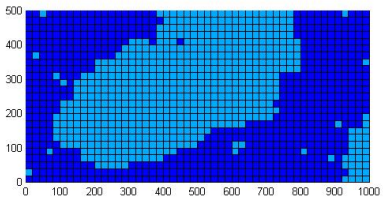
Figure 7.1: Depicted are the sub-communities with the highest mean relative density for each 20-by-20 m quadrant for the number of cluster k between $k = 6$ and $k = 10$ (top-left to bottom-right) for the adult plants in the 2010 census.

7.4 Sub-community maps for 5 and 20 metres scale

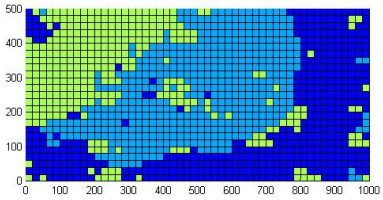
The sub-community maps at 5 m, 10 m and 20 m show similar structures. At the 5 m scale the map looks somewhat less smooth. At this scale the method takes into account small scale structuring factors such as canopy gaps more strongly, while environmental variables such as the topography that structures the area have a larger scale influence. The swamp habitat can be distinguished at the 20 m scale already from $k = 4$ clusters, while it only forms a separate cluster from $k = 5$ at the 10 m scale, and only from $k = 6$ at the 5 m scale. But the main distinctions (slope, high plateau, south-east low plateau, north-west low plateau, and swamp) can all be found within the first six clusters at all scales.



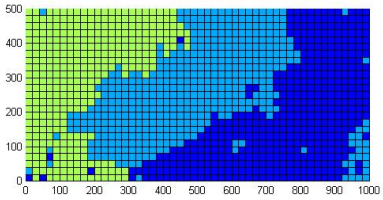
(a) k = 2; 5 m



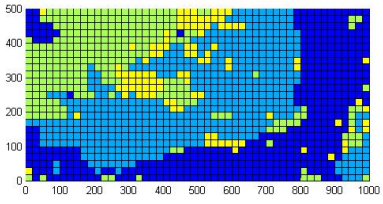
(b) k = 2; 20 m



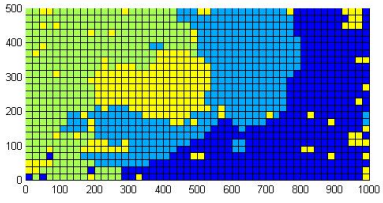
(c) k = 3; 5 m



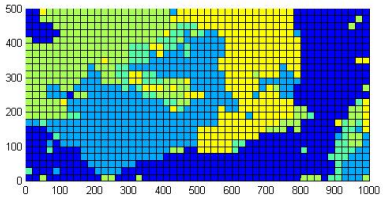
(d) k = 3; 20 m



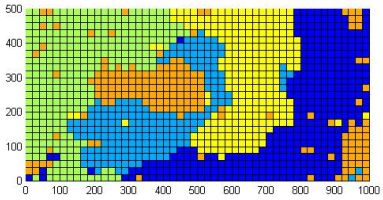
(e) k = 4; 5 m



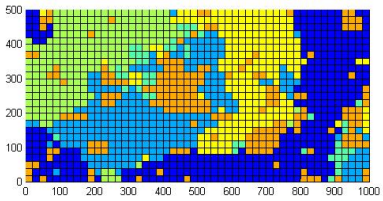
(f) k = 4; 20 m



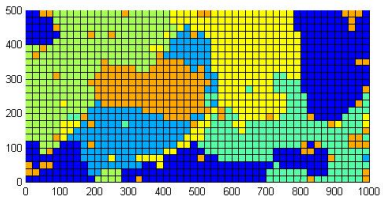
(g) k = 5; 5 m



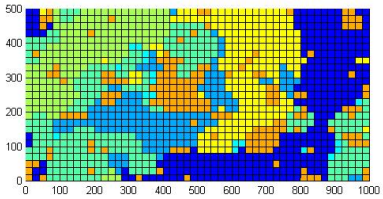
(h) k = 5; 20 m



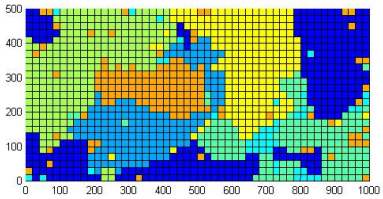
(i) k = 6; 5 m



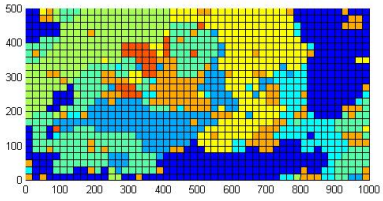
(j) k = 6; 20 m



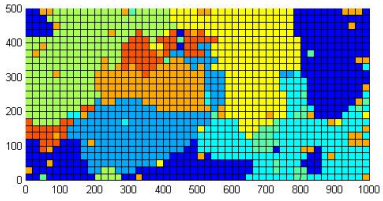
(k) k = 7; 5 m



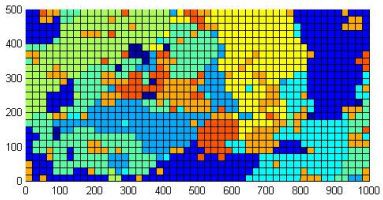
(l) k = 7; 20 m



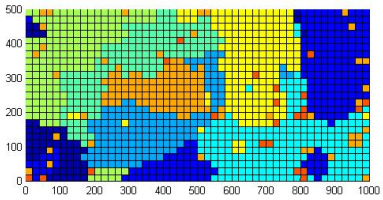
(m) k = 8; 5 m



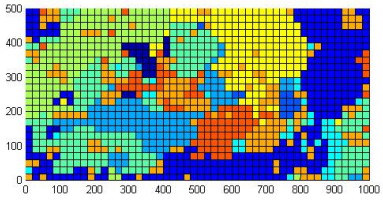
(n) k = 8; 20 m



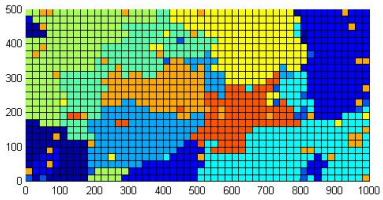
(o) k = 9; 5 m



(p) k = 9; 20 m



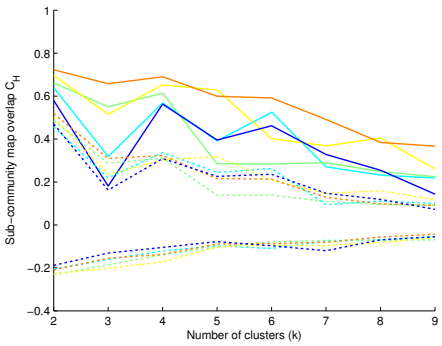
(q) k = 10; 5 m



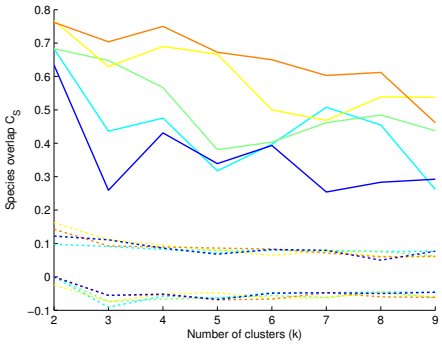
(r) k = 10; 20 m

Figure 7.2: Depicted are the sub-communities with the highest mean relative density for each 20-by-20 m quadrant for the adult plants in the 2010 census; Left column with a neighbourhood radius of 5 metres, right column with a radius of 20 metres.

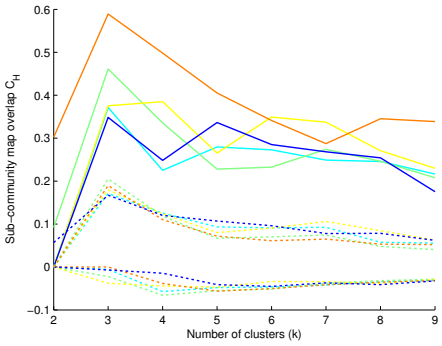
7.5 Significance of sub-community stability over-time



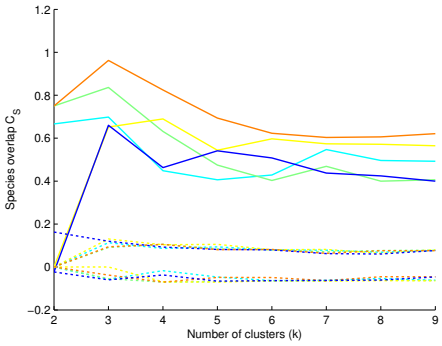
(a) Stability of sub-community maps in adults.



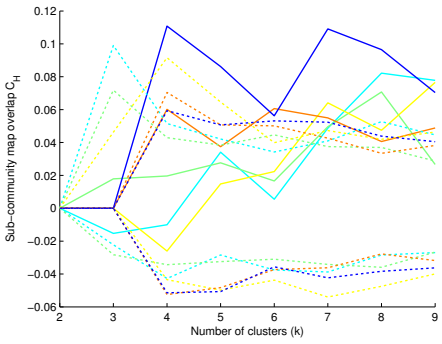
(b) Stability of species composition in adults.



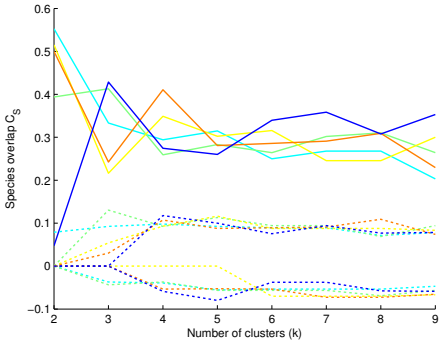
(c) Stability of sub-community maps in juveniles.



(d) Stability of species composition in juveniles.



(e) Stability of sub-community maps in recruits.



(f) Stability of species composition in recruits.

Figure 7.3: The stability of the sub-communities between 1985 and 2010. Depicted are the consistencies between the sub-community map derived from the 2010 census data and earlier time points (left column) and the consistencies between species compositions of sub-communites (right column). Solid lines show results for the Barro Colorado data, and dotted lines show the 2.5% and 97.5% percentile of the stability of 1000 randomised sub-community maps (left column) or 1000 randomised species assignments to sub-communities (right column) based on the same data. In panels (a) and (b) we show the results for the adults, panels (c) and (d) show the results for juveniles and panels (e) and (f) those for the recruits. The different colors show different censuses which we compared to the 2010 census: 1985 (dark blue), 1990 (light blue), 1995 (green), 2000 (yellow), 2005 (orange).